CONSERVATION GENETICS OF AFRICAN ELEPHANTS

BY

ALIDA DE FLAMINGH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Ecology, Evolution and Conservation Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Doctoral Committee:

Professor Alfred L. Roca, Chair and Director of Research
Professor Ripan S. Malhi
Professor Robert L. Schooley
Professor Rudi J. van Aarde, University of Pretoria

ABSTRACT

Elephant population numbers are plummeting across the African continent. Habitat loss as a consequence of anthropogenic landscape transformation, and the poaching of elephants for the illegal ivory trade, has reduced and fragmented elephant populations to a fraction of their former population size and range. It has therefore become necessary to develop and implement proactive and targeted conservation initiatives that aim at maintaining or restoring spatial and genetic connectivity between populations. In this thesis, I use a conservation genetic approach to study African elephant populations with the aim of developing methodologies and providing applications that could inform conservation planning for African elephants.

In Chapter 1, I introduce and discuss the need for proactive African elephant conservation initiatives. In Chapter 2, I develop and evaluate an approach to delineate functional landscape linkages (FLL) for African elephant conservation planning, and I use this approach to create a fine-scale map that demarcates FLL for elephant populations in southern Africa. The results of Chapter 3 provide support for conservation initiatives that aim at increasing connectivity through FLL. In Chapter 3, I show that Kruger National Park's elephant population forms part of a functional entity in which migration helped to maintain a relatively diverse gene pool. Chapter 4 benefits elephant conservation by increasing our understanding of historical elephant population ecology and genetics. Here I show that contemporary West African forest elephant populations have limited genetic diversity compared to the genetic diversity found in historical populations. Chapter 4 highlights the need for proactive and preventative conservation strategies that aim to conserve the genetic diversity within remaining forest elephant populations. In Chapter 5 I adapt a method previously developed for sex identification of human remains for use with non-human taxa, and I successfully identify the sex of modern and ancient elephants from low coverage

genome data. Sex identification of ancient animal biological remains can benefit conservation by increasing our understanding of historical population structure, demography and social behavior.

In this thesis I develop and apply genetic and spatial analyses to extinct and extant elephant populations to inform local and regional African elephant conservation strategies. This thesis provides an interdisciplinary toolset and framework for future conservation genetic studies that focus on African elephant conservation planning.

ACKNOWLEDGEMENTS

This thesis could not have been completed without the support and help of many friends and colleagues, only some of whom it is possible to give particular mention here. Foremost, I would like to express my sincere gratitude to my doctoral advisor, Dr. Alfred Roca, for his support and trust, and for the many interesting opportunities that working with him presented. Thanks, Al, for all of the intellectually stimulating discussions, for supporting many of my academic endeavors, for your patience and for always being willing to share your immense knowledge with me. I would also like to thank Dr. Ripan Malhi for his support, encouragement and guidance. Ripan, thank you so much for broadening my academic (and personal) horizons, and for encouraging me to pursue research that I never knew I would be able to do and love, and thank you very much for the wonderful research trips to Alaska.

My sincere thanks to my other doctoral committee members, Dr. Robert Schooley and Prof. Rudi van Aarde. Your advice and knowledge provided much needed guidance throughout my doctoral studies. I am also very grateful to my research program, the Program in Ecology, Evolution and Conservation Biology (PEEC) for the continuous financial and academic support during my PhD.

Thank you to my fellow Roca and Malhi lab members and the Graduates in Ecology and Evolutionary Biology (GEEB), for all of the stimulating discussions and fun adventures. A special thanks to Tolu Perrin-Stowe and Angel Rivera-Colón for always having my back.

Last but definitely not least, thank you to my family and friends, especially those in South Africa and elsewhere abroad, who kept me grounded and made me feel loved regardless of the oceans between us.

I dedicate this thesis to my wonderful parents, Louisa and Jan de Flamingh, who wholeheartedly supported me through it all, and who early-on inspired in me a love and respect for nature through many adventures into the beautiful wilderness of southern Africa. I also dedicate this thesis to Tilly and Willem Buys, for their unwavering support and encouragement, and for being my best friends. Ek is ontsettend lief vir elkeen van julle.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

In 2016, the African Elephant Specialist Group reported a continental decline in elephant population numbers. In the previous decade, populations decreased by more than 100 000 elephants due to illegal poaching for ivory (CITES, 2016), and elephant populations continue to decrease at an alarming rate (Thouless *et al.* 2016). Their distributional range is now a fraction of what it once was (Ripple *et al.* 2015). In Africa alone, protected areas are estimated to contain only a quarter of the expected number of elephants (Robson *et al.* 2017). With a global increase in poaching (Wittemyer *et al.* 2014), proactive conservation efforts have become a necessity. In this thesis I use a conservation genetic approach to study African elephant populations to develop methodologies and provide applications that could inform African elephant conservation planning.

The field of conservation genetics aims to uncover information about populations that could help conserve them (Beaumont and Wang 2019). Applications of conservation genetics analyses include, for example, estimating the effective population size and census population size (Waples 2016; Murphy *et al.* 2018; Pelletier *et al.* 2019), assessing the deleterious effects of inbreeding and the possibility of genetic rescue (Hedrick and Garcia-Dorado 2016; Seddon and Schultz 2020), elucidating population structure and phylogeography (de Flamingh *et al.* 2017; Murphy *et al.* 2018; Pelletier *et al.* 2019), inferring gene flow, migration and estimating population connectivity (Hoban 2018; Monti *et al.* 2018; Cameron *et al.* 2019). Conservation genetics therefore provides tools that can be used to develop efficient and effective conservation strategies to protect and manage wildlife populations.

In this thesis, I apply conservation genetic analyses to African savanna elephants (*Loxodonta africana*) in southern Africa (Chapters 2, 3 and 5), and to "historic ivory" from West

African forest elephants (*Loxodonta cyclotis*) (Chapters 4 and 5). These two African elephant species differ morphologically (Groves 2000; Groves and Grubb 2000; Grubb *et al.* 2000) and genetically (Roca *et al.* 2001; Ishida *et al.* 2011). Female elephants of both African elephant species form fission-fusion societies (Fishlock and Lee 2013) and are matrilocal, remaining with their natal social groups throughout their lives in a relatively restricted geographic range, whereas dispersal in elephants is male-mediated (Archie *et al.* 2007; Fishlock and Lee 2013). Mitochondrial DNA (mtDNA), which follows a matrilineal inheritance pattern, may therefore allow for the determination of the geographic provenance of elephants (Ishida *et al.* 2013). However, ancient hybridization and backcrossing may have resulted in the transfer of African forest elephant mtDNA to African savanna elephants, and as a consequence species identification should consider nuclear DNA (Roca *et al.* 2015).

I investigate the genetics of these African elephant species across varying temporal and spatial scales. The spatial scale of this thesis includes regional elephant populations spanning 7 different countries in southern Africa (Chapter 2), local elephant populations within the Kruger National Park (Chapter 3), DNA from elephants with unknown provenance (Chapter 4), and a methodological approach to identify the sex of elephants regardless of geographic origin or species (Chapter 5). The temporal scale of this thesis includes data from contemporary elephant populations (Chapters 2 and 3), and ancient ivory DNA from a ~500-year-old shipwreck (Chapters 4 and 5).

I develop and implement various analytical approaches to answer critical questions related to African elephant conservation. These chapters are linked in theme because they each provide information or methodological tools that could benefit elephant conservation science through the application of novel scientific techniques, and by promoting the advancement of

integrative approaches to deal with conservation problems. In Chapter 2, I identify functional landscape linkages (FLL) by integrating spatial (GPS telemetry) data and genetic data for elephant populations in southern Africa. Functional landscape linkages (FLL) are wildlife migration corridors that increase connectedness between isolated or fragmented populations, and in so doing increase gene flow and genetic diversity, minimizing the genetic isolation and inbreeding that can occur in unconnected protected areas. Thus, FLL can increase the genetic health and the persistence of wildlife populations (Allendorf *et al.* 2013). In Chapter 3, I investigate the geographic origin of elephants in Kruger National Park (South Africa), and examine regional phylogeographic patterns by comparing the mtDNA and nDNA characteristics of elephants in Kruger National Park to populations in nearby countries (Botswana, Mozambique, Zambia and Zimbabwe). In Chapter 4, I use paleogenomic and stable isotope approaches to analyze 500-year old elephant tusks recovered from a 16th century shipwreck, providing information on the species of elephant, their geographic provenance, and habitat characteristics of the landscape in which they lived. With a resolution not possible using any single approach, combining paleogenomic, stable isotope, and archeological methods allows for the examination of patterns of ivory acquisition and circulation during the formative stages of maritime trade that linked Europe, Africa and Asia. In Chapter 5, I adapt a method developed for human sex determination so that it can be used to identify the sex of other ancient and modern animal taxa. I test the accuracy of this method using low coverage genomes from 15 modern savanna elephants for which sex was known, and then applied this method to the shipwreck ivory samples for which sex was unknown. Because this method is amenable to low coverage data from low quantity DNA (e.g., ancient or degraded DNA), it can be employed as a non-invasive approach to identifying sex of endangered or rare species, for example, through the analysis of

DNA from hair tufts (McKelvey *et al.* 2006; Stanton *et al.* 2016) or scat (Huber *et al.* 2002). This method can also be used to identify the sex of organisms in which the reference genome was obtained from a female animal or where only the X and not the Y chromosome is reported in the reference genome assembly for the taxon. Such taxa would include (but not be limited to) the domestic cat, sheep, goat, horse, dromedary camel, European rabbit; and also include many wild animals such as the Sumatran orangutan, western lowland gorilla, gelada and meerkat.

Collectively these chapters provide interdisciplinary tools for African elephant conservation planning. I develop and apply genetic and spatial analyses to extinct and extant elephant populations to inform local and regional conservation strategies. My thesis provides a framework for future studies that aim to address elephant conservation issues by using a conservation genetics approach.

CHAPTER 2: FUNCTIONAL LANDSCAPE LINKAGES FOR AFRICAN SAVANNAH ELEPHANT (*LOXODONTA AFRICANA*) CONSERVATION PLANNING

2.1 ABSTRACT

Space for conservation is becoming an increasingly limited resource as human densities rise, land is transformed and natural habitat is destroyed. Targeted conservation initiatives such as the demarcation of functional landscape linkages (FLL) as wildlife movement corridors may be needed to increase or maintain dispersal, gene flow and genetic diversity between and within isolated populations. Gene flow counteracts the effects of genetic drift, inbreeding, expression of deleterious alleles, and increased extinction risk often associated with isolated populations. In this study, we identify FLL between African savannah elephant (*Loxodonta africana*) populations that occur across 7 southern African countries by combining habitat suitability modelling with gene flow patterns. We identify areas in the landscape that provide for the habitat requirements of elephants and that also show evidence of gene flow. We evaluate different habitat suitability models (HSM) predicted using MaxEnt and transform the best performing HSM to resistance surfaces that allowed for linear, slight nonlinear and pronounced nonlinear responses that elephants may have to unsuitable habitats. We evaluate these resistance surfaces using maximum likelihood population effects (MLPE) models to determine whether geographic Euclidean distance, least-cost path or multipath distances in circuit theory best explain the observed pattern of gene flow across our study area. Based on the MLPE results, we used the pronounced nonlinear resistance surface in combination with genetic distances to produce an FLL map for elephants in southern Africa. Our final FLL map considers suitable habitats based on spatial data, and also gene flow as the inverse of genetic distance, to delineate areas in the southern African landscape that may be important for maintaining or restoring elephant

population connectivity. This study develops and applies a novel conservation genetic approach that integrates multidisciplinary data and methods to inform conservation planning. The approach developed here may be used as a framework for studies on other taxa that aim to develop spatially and genetically informed conservation strategies.

2.2 INTRODUCTION

Space for conservation is becoming an increasingly limited resource as human densities rise, land is transformed and natural habitat is destroyed. In Africa, protected areas have become progressively more isolated (Newmark 2008; Ripple *et al.* 2015) and conservation initiatives may need to focus on increasing or maintaining population connectivity to support population persistence through increased dispersal, gene flow, genetic diversity and environmental adaptability (Garant *et al.* 2007).

Functional landscape linkages (FLL) are wildlife migration corridors that increase the connectedness between isolated or fragmented populations, and in so doing increase gene flow and genetic diversity, minimizing the genetic isolation and inbreeding that can occur in unconnected protected areas. Thus, FLL can increase the genetic health and the persistence of wildlife populations (Allendorf *et al.* 2013). This study sought to identify FLL by integrating spatial data from GPS-marked individuals and genetic data for elephant (*Loxodonta africana*) populations in southern Africa.

FLL may be especially beneficial to elephant populations because their current distribution across much of Africa is patchy and greatly affected by habitat loss (van Aarde and Jackson 2007; Graham *et al.* 2009), obstruction of dispersal opportunities (Loarie *et al.* 2009), and extirpation of local populations due to hunting and ivory poaching (Whitehouse and Harley 2001; Okello *et al.* 2008; Thouless *et al.* 2016). These forces have reduced local population sizes

and compressed remnant populations into protected areas (Laws 1970). Many elephant populations, even when protected, exist as discontinuous units. Moreover, Robson et al. (2017) recently estimated that, due to poaching, only a quarter of the expected number of elephants currently occur in protected areas across Africa.

FLL are identified by evaluating areas of suitable habitat using spatial data, and by examining genetic data to identify the gene flow that has occurred between populations in these areas. Landscape linkages are most robust when identified by integrating these two sources of information. Spatial data may demonstrate movement and dispersal patterns, but by themselves will not provide evidence for successful migration and reproduction (gene flow). By contrast, genetic data may establish that migration and gene flow have occurred between two populations, but do not identify specific routes that need to be conserved to maintain movement between populations. In this study, we sought to generate a comprehensive, empirically evaluated map of functional landscape linkages based on spatial movement and genetic data for elephants across southern Africa.

We used telemetry data from 116 elephants in southern Africa, in combination with data from four environmental variables that can influence elephant space use, to predict areas of suitable habitat that may provide for the habitat requirements of elephants, i.e., the potential distribution of elephants. We predicted their potential distribution using a maximum-entropy (MaxEnt) modelling approach (Phillips *et al.* 2017). MaxEnt estimates the potential distribution of animals by finding the spatial distribution of maximum entropy (the distribution that is closest to being uniform) when comparing the expected value of each environmental variable under this estimated distribution to the empirical average that has been calculated from the occurrence or telemetry data (Merow *et al.* 2013). We assessed various MaxEnt model optimization parameters

(for example, we tested various regularization parameters to decrease model over-fitting) and also tested three different sets of background datasets to which presence locations were compared in MaxEnt. We used the MaxEnt predicted potential distribution for the highest scoring model based on the area under the receiver-operator curve (AUC) values for 5-fold cross-validated models to create a single best habitat suitability map (HSM) for elephants. We then transformed this HSM linearly and nonlinearly through a negative exponential function into resistance surfaces (Keeley *et al.* 2016). Resistance is usually assumed to be a negative linear function of suitability (Hunter *et al.* 2003; Larkin *et al.* 2004; Pullinger and Johnson 2010). However, in addition to a negative linear transformation, we also tested slight and pronounced nonlinear transformations to investigate nonlinear responses that elephants might have to unsuitable habitats in the resistance surfaces. Landscape resistance surfaces are grid maps that represent the resistance to movement through the landscape features contained in that grid cell (Milanesi *et al.* 2017), or the cost of moving through such a landscape.

We calculated three different measures of connectivity across each transformed resistance surface, including standard Euclidean distance, least-cost path (LCP) distance (Cushman *et al.* 2006) and circuit theory (CT) resistance distance (McRae and Beier 2007). These measures of connectivity were calculated in a pairwise fashion across 142 elephant sample locations for which we had microsatellite data for 9 nuclear DNA loci.

The metric of connectivity (Euclidean distance, LCP, CT) that best described the genetic patterns observed across the landscape was identified using Akaike's information criterion, marginal R^2 and conditional R^2 values calculated for maximum likelihood population effects (MLPE) models. We fitted MLPE models between pairwise matrices of genetic distances and matrices of Euclidean, LCP or CT distances. Euclidean, LCP and CT distances represent the

fixed effects in each model while the genetic distances among individuals represent the response variable. MLPE models also include pairwise random effects for each distance comparison to account for sample autocorrelation and multiple pairwise distances per site (Row *et al.* 2017).

Our MLPE model assessment identified CT distances as the connectivity measure that best explained the observed gene glow across our landscape. We therefore used CT distances calculated in Circuitscape (Shah and McRae 2008) to estimate, delineate and visualize areas in the landscape that represent FLL. Our final FLL map considers suitable habitats based on spatial data, and also gene flow as the inverse of genetic distance, to delineate areas in the landscape that may be important for maintaining or restoring population connectivity.

2.3 METHODS

2.3.1 Study area and elephant occurrence data

Our study area spans seven countries across southern Africa, including Botswana, Malawi, Mozambique, Namibia, South Africa, Zambia and Zimbabwe (Figure 2.1), and includes a range of different vegetation classes (Figure A.1). Collectively these countries contain >70% of the elephants in Africa and >42% of the total range of elephants in Africa (Thouless *et al.* 2016).

Through a collaboration with the Conservation Ecology Research Unit (University of Pretoria, South Africa) we had access to an extensive database of telemetry (GPS collar) locations for 116 elephants (Figure 2.1 – Elephant occurrence data). Of these 116 elephants, 95 were from elephant breeding herds and 21 were male elephants. We used quality filtering criteria similar to that of Roever, van Aarde and Leggett (2013) to select reliable telemetry locations. To decrease temporal and spatial autocorrelation, we filtered the data so that only a single point per day per elephant was retained. Our spatial database included a total of 81 858 location points. The temporal scale of data varies per individual since GPS collars were deployed in different

years, but overall the elephant monitoring occurred from 2003 to 2015 across all seasons. These elephants form part of 6 regional population clusters that include the Chobe, Kafue, Limpopo, Luangwa, Niassa and Zambezi clusters (Figure A.2).

2.3.2 Elephant genetic data

Fresh elephant fecal samples were collected from six southern African countries from 2010-2014 (Figure 2.1 – Elephant genetic samples). A total of 142 samples were collected and genotyped for 9 highly variable nuclear DNA microsatellite loci (de Flamingh, Roca and van Aarde, 2018; Table A.1). Sample collection, DNA extraction procedures, and microsatellite amplification are detailed in de Flamingh, Roca and van Aarde (2018) and genotype errors were quantified as described in de Flamingh, Sole and van Aarde (2014). Heterozygous genotypes were replicated at least four times, and homozygous genotypes were replicated at least three times. Diversity indices for individual microsatellite loci were calculated in Arlequin (Excoffier and Lischer 2010).

We assumed an inverse relationship between gene flow and genetic distance (GD), where areas of high gene flow would result in low genetic distances among individuals and *vice versa*. Different estimates of GD vary in their ability to capture variation at the landscape scale (Shirk *et al.* 2012). Using the program R (R Core Team 2019), we therefore quantified pairwise GD using four alternative GD estimates that have regularly been used in landscape genetic studies (Shirk *et al.* 2012; Kamvar *et al.* 2014; Milanesi *et al.* 2017; Tang *et al.* 2019). Some GD estimates rely on the same data variables (e.g. allele presence and frequency) to quantify GD. The GD measures presented in this study are therefore not independent and there may be overlap between GD quantifications and model outcomes when GD estimates with highly similar data variables are used as response variables for model fitting. To represent response variables in our MLPE

models, we calculated GD as 1) a value of 1 minus the proportion of shared alleles (D_{PS}) using the "propShared" function in the R package "adegenet" (Jombart 2008); 2) the number of allelic differences between two individuals using the "diss.dist" function in the R package "poppr" (Kamvar *et al.* 2014); 3) the Euclidean distance among a vector of allele frequencies using the "dist" function in the R package "adegenet"; 4) Reynolds's distance with the "Reynolds.dist" function in the R package "poppr".

To visualize gene flow across the landscape, we interpolated genetic distances to form a landscape shape in the program *Alleles in Space* (AIS; Miller, 2005). AIS creates a connectivity network between all sample locations and places genetic distances as midpoints of each pairwise connection. The program then interpolates genetic distances across the extent of the study area and produces a 3-dimensional surface plot where surface heights represent interpolated genetic distances, and where higher peaks in the surface plot indicate greater genetic distances. We tested for correlation between genetic and geographic distances, and calculated GD in AIS using a GD estimate identical to that of Nei, Tajima and Tateno (1983). We used a distance weighting parameter (*a*) of 1 to interpolate genetic distance across the landscape (Miller 2005). We exported the interpolated output using the highest possible resolution of 500 x 500 bins for the X and Y geographic axes, and plotted for each of the bin coordinates the peak heights using the program ArcMap V10.7.1 (© ESRI 2011).

2.3.3 Environmental variables

We considered habitat covariates that are known to influence elephant space use. These environmental variables included water availability (Loarie *et al.* 2009), the slope of the land (Wall *et al.* 2006), primary productivity (Young *et al.* 2009b) and human presence (Hoare and Du Toit 1999).

2.3.3.1 Water availability

Surface-water availability drives the distribution and abundance of elephants (Chamaillé-Jammes *et al.* 2007). We used the Global Surface Water (GSW) occurrence layer (Pekel *et al.* 2016) to estimate water availability across our study range. The GSW occurrence layer represents areas where surface water occurred between 1984 and 2018 and provides information concerning overall water dynamics (e.g., intra and inter-annual variability and change). This layer captures the frequency with which water was present in a given area. To calculate GSW occurrence, Pekel *et al.* (2016) summed and normalized monthly water detections (WD) and valid observations (VO) such that GSW occurrence/month = ∑WD month / ∑VO month. By averaging the results of all monthly GSW occurrence calculations Pekel *et al.* (2016) were able to provide the long-term overall surface water occurrence.

GSW data were downloaded as 9 individual tiles that covered our study extent and were merged using the "Mosaic to new raster" function in ArcMap V10.7.1 (© ESRI 2011). We transformed the GSW data to represent the geographic distance from an available water source since areas that are further from water sources would represent less suitable habitat, and a gradient layer may therefore be more biologically relevant to elephant space use than a binary presence-absence layer. We reclassified the GSW data so that all points with values greater than 0 were reclassed to represent water occurrence. We used the Euclidean distance tool in ArcMap to calculate, for each cell in our study area, the Euclidean distance to the closest water source. The GSW dataset was split into subset of four raster layers before executing the Euclidean distance tool to facilitate data processing. The four distance layers were merged into a single layer which was used as one of four input environmental layers in MaxEnt (Figure 2.2B).

2.3.3.2 Slope

Elephants alter their space use to avoid mountainous terrain where even small hills may present energy barriers to movement for large bodied animals (Wall *et al.* 2006). Here we convert a digital elevation model (Jarvis *et al.* 2008) to denote slope by calculating the maximum rate of elevation change between pixels using the "Spatial Analyst" toolbox in ArcMap (ESRI © 2011). We use this converted slope layer as our second environmental layer in MaxEnt (Figure 2.2C).

2.3.3.3 Primary productivity

Primary productivity as a measure of food availability influences elephant habitat selection and use of space (Young *et al.* 2009a; Roever *et al.* 2012). As our third environmental variable in MaxEnt, we used long-term mean Enhanced Vegetation Index (EVI) from Robson *et al.* (2017), which is an index of primary productivity across our study range as (Figure 2.2D). We use EVI rather than Normalized Difference Vegetation Index (NDVI) because EVI overcomes some of the contamination problems present in NDVI data (e.g., contamination associated with canopy background and residual aerosol influences) and is less likely to become saturated in areas that have high green biomass (Pettorelli *et al.* 2005).

2.3.3.4 Human presence data

Elephants alter their space use to avoid densely populated areas (Barnes *et al.* 1991; Hoare and Du Toit 1999), and we therefore included human presence as the fourth environmental variable in our MaxEnt models. We used Landscan2016 global population distribution data, which is the finest resolution human population data available. Landscan data represents an "ambient population" (average population presence over 24 hours; Bright *et al.*, 2017). LandScan data are preferable to other census based data because it accounts for differences in spatial data availability, scale and accuracy, and Landscan distribution models are

specific to individual countries and regions (Bright *et al.* 2017). Furthermore, LandScan data as an "ambient population" rather than point density considers both diurnal movements and collective travel habits, and integrates these variables into a single measure (Dobson *et al.* 2000).

Other studies have transformed human density data to a gradient-based distance metric (Figure 2.2E) that represents geographic distances from densely populated areas (Roever *et al.* 2013). We evaluated transformed gradient-based human density but found that it performs poorly compared to raw LandScan "ambient population" data in our MaxEnt models. See Table A.2 for the area under the receiver-operator curve (AUC) results of gradient-based human presence MaxEnt models. In addition, the transformed gradient-based human distance metric was also correlated with our GSW environmental variable ($r > 0.60$; Figure A.3). We therefore used raw ambient human population data (Figure 2.2F) rather than transformed gradient-based distance data (Figure 2.2E) to represent human presence in our MaxEnt models. Using raw ambient human population data may also be better aligned with the nonlinear relationship where elephant and human coexistence occurs at a range of human densities and local elephant population absences are contingent on landscape-dependent thresholds of human density (Hoare and Du Toit 1999). Hoare (1999) showed that male elephants are more likely to come into conflict with humans (e.g. raid crops) and may therefore be less concerned about moving through densely populated areas than female elephants. In light of these sex-based behavioral differences, we discuss the possibility of delineating sex-specific FLL in the "future directions" section.

2.3.3.5 Raster preparation for MaxEnt modelling

Maxent requires all environmental layers to have the same resolution and extent. We resampled environmental layers to 900m X 900m (0.0083 x 0.0083 decimal degrees; standardized across environmental data layers) so that resolution represented a biologically

relevant scale in view of elephant space use and distribution. The spatial resolution of our layers are informative for a highly mobile species such as elephants (Young, Ferreira and van Aarde, 2009) but are still much smaller than average home range size of elephants in these population clusters (Grainger, Van Aarde and Whyte (2005); Roever, Van Aarde and Leggett (2012); please see the "Future directions" section on our planned landscape layer up- and down-scaling analysis). We used a bilinear interpolation approach that is suitable for continuous data to resample our environmental layers using the "Data management" toolbox in ArcMap (ESRI © 2011). Bilinear interpolation calculates the value of each pixel by averaging the values of surrounding pixels. We standardized the coordinate system and datum to WGS_1984 for all input layers, and tested for correlation among our environmental variables by calculating Pearson's correlation coefficient in the R package "virtualspecies" (Leroy *et al.* 2016; R Core Team 2019).

2.3.4 MaxEnt Habitat Suitability Modelling

We used a MaxEnt modelling approach to determine which environments may represent suitable habitats for elephants, and subsequently predicted where such suitable habitats may occur across our study area. MaxEnt requires presence-only data to model species distribution and predict habitat suitability (Elith *et al.* 2011). MaxEnt uses occurrence data points (presences) in combination with a set of environmental predictor variables (e.g., GSW, slope, EVI, human presence), and compares these presences and their associated environmental constraints to a set of background points at which presence is unknown (Merow *et al.* 2013). MaxEnt then predicts the relative occurrence rate (ROR) for each cell in the landscape grid, where ROR is the relative probability that the cell is contained in the presence data, and where ROR is contingent on selecting the "raw" output option in MaxEnt (Elith et al 2011).

MaxEnt can also be used to estimate the probability of presence by using a logistic transformation of the ROR. Logistic probability of presences may improve model calibration and are contingent on selecting the "logistic" output option in MaxEnt (Phillips and Dudík 2008). Elith *et al.* (2011) and Royle *et al.* (2012) warn against using logistic rather than raw data because the variation in predictive outcomes are highly dependent on the assumptions made when MaxEnt logistically transforms the data *ad hoc*. However, the "raw" output is difficult to interpret, especially when large background datasets such as ours are used (see "Background data" section) and where scale-dependence may result in small "raw" values (Phillips and Dudík 2008). Also, we will only be comparing map predictions that were generated using the same assumptions to estimate the probability of presence, and will not be comparing surfaces that were based on, for example, different values of tau (τ; Merow, Smith and Silander Jr, 2013). Here we use the "logistic" rather than "raw" output option in MaxEnt to allow for large differences in output values to translate to large differences in suitability. As expected, the large value ranges provided by the logistic transformation performed better than "raw" output when linear and nonlinear transformations were used to convert HSM to resistance surfaces. The "raw" output resulted in minimal differences in resistance surfaces, while "logistic" output resulted in resistance surfaces that were observably different.

2.3.4.1 MaxEnt background data

Choosing relevant background location data for comparison to presence data is critical when the MaxEnt model will be used to predict or extrapolate to novel environments where presence data are not available (Elith *et al.* 2011; Webber *et al.* 2011). Background data should

be chosen to reflect the environmental conditions that are relevant to the species for which the model is generated, and they should be based on the spatial scale of the ecological questions of interest (Saupe et al 2012, Merow et al 2013). AUC can be inflated by increasing the number of background points, and by selecting background data that may not be ecologically informative for species distribution modelling (Lobo et al. 2008, Anderson 2012).

We aimed to delineate functional landscape linkages within and between areas where elephants are known to occur. We were also interested in delineating linkages that may not currently fall within the known elephant range, but where future conservation initiatives may aim to conserve these linkages in effort to re-establish connectivity between isolated populations. We therefore tested three different background datasets: background data generated for the entire extent of all environmental predictors included in the model (Figure 2.3B); range-based background data generated for areas of known elephant range as demarcated by the IUCN Red List of Threatened Species (Figure 2.3C, Blanc 2008); and use-based background data from areas included in an 80% convex hull of the presence points (Figure 2.3D). For each of these datasets we generated 10 000 random points to compare to the presence data.

2.3.4.2 MaxEnt model regularization

Model regularization can reduce model over-fitting by decreasing model complexity (Merow *et al.* 2013). Regularization may allow for less precise fitting of the empirical constraints from environmental features in the MaxEnt model, and it simplifies models by incorporating a penalty that is proportional to the magnitude of the regularization coefficient (Merow et al 2013). Therefore, using explicit regularization parameters may prevent model complexity from increasing beyond what is supported by the empirical dataset (Phillips and Dudík 2008). We tested four regularization beta parameters in Maxent ($\beta = 1, 2, 3, 4$), and compared AUC values

for cross-validated models as a measure of model fit to determine the most suitable beta parameter for our dataset. Regularization parameters were tested for all 3 background datasets

2.3.5 Model evaluation

We evaluated the Maxent models using k-fold cross validation using a k of 5. We withheld a random sample of 20% of the presence data for testing and trained our model with the remaining 80% of the data. We assessed the evaluated models by investigating the area under the receiver-operator curve (AUC). AUC calculated the probability that a presence location ranks higher than a random background point in the predicted model (Merow *et al.* 2013). We compared AUC values for three types of background data, where each type of background data was fitted using $\beta = 1, 2, 3, 4$ (12 models in total). We selected the best scoring combination of background type and regularization parameter to generate our HSM.

2.3.6 Resistance surfaces and connectivity metrics

The single best HSM was selected and transformed linearly and nonlinearly into resistance surfaces (Keeley *et al.* 2016). A linear transformation assumes that cost (R_{lin}) is inversely related to habitat suitability, where less suitable habitat would be costlier to move through (Eq. 2.1).

$$R_{lin} = 1 - HSM \qquad \text{Eq. 2.1}$$

R_{lin} is the landscape resistance based on a linear transformation of habitat suitability (HSM) (Figure 2.4A). A nonlinear transformation by means of a negative exponential function allowed for the possibility of nonlinear responses that elephants might have to unsuitable habitats. A nonlinear response would be, for example, when elephant movement across the landscape is only substantially impacted by highly unsuitable habitats (e.g., very high cost areas), but where

medium- and low-cost areas do not impact movement to the same effect. We used a negative exponential function to nonlinearly transform HSM to cost, R (Eq. 2.2).

$$R = 100 - 99 * ((1 - \exp(-c * H))/(1-\exp(-c))) \qquad \text{Eq 2.2}$$

where R is resistance, H is suitability, and the factor c determines the shape of the curves. We generated resistance surfaces for c = 2 (slight nonlinear transformation) and c = 8 (pronounced nonlinear transformation) to allow for variable strengths of nonlinear responses (Figure 2.4B).

2.3.7 Spatial and genetic connectivity across landscape resistance surfaces

For each of the transformed resistance surfaces, we calculated three commonly used measures of landscape connectivity (McRae and Beier 2007) in a pairwise fashion between 142 elephant sample locations for which we had microsatellite data for 9 nuclear DNA loci (Table A.1). These measures of connectivity included standard geographic Euclidean distance, Least-Cost Path (LCP; Cushman *et al.*, 2006), and resistance distance based on circuit theory (CT; McRae and Beier, 2007). These pairwise distance matrices were used as fixed effect predictor variables in the MLPE models.

2.3.8 Maximum likelihood population effects (MLPE) model comparison

The measure of connectivity that best described the genetic patterns observed across the landscape was identified using Akaike's information criterion calculated in the "lme4" package in R (Burnham and Anderson 2004; Bates *et al.* 2015), and using marginal and conditional R^2 values calculated in the R package "MuMIn" (Barton 2009). The lowest AIC values indicate the MLPE model with the best fit (Burnham and Anderson 2004), higher marginal R^2 values represent higher predictive power of the fixed effects, and higher conditional R^2 represents higher total variance explained by the fixed effects (Edwards *et al.* 2008). Compared to traditional R^2 values, marginal and conditional R^2 values are favorable model selection criteria

(Van Strien *et al.* 2012) since they do not necessarily increase with the addition of model parameters (Orelien and Edwards 2008). Restricted maximum likelihood (REML) estimation of MLPE models has been suggested for determining accurate estimators of R^2 values (Verbeke and Molenberghs 2000; Gurka 2006), and we therefore calculated conditional and marginal R^2 values using REML. However, REML should not be used when comparing information criteria for MLPE models with different fixed effects, and we therefore did not include REML when we calculated AIC.

Euclidean, LCP and CT distances represent the fixed effects in each model while the four quantifications of pairwise genetic distances among individuals represent the response variables. Using the "base" package in R (R Core Team 2019), we scaled our response variables prior to fitting MLPE models to allow for the comparison with predictor variable values. The scale function in "base" centers and scales the columns of a numeric matrix (Becker *et al.* 1988). MLPE models account for the nonindependent structure associated with pairwise distance matrices by considering as random effects a population-level factor that identifies data points that share a common "deme", i.e., corresponding data points in the pairwise distance matrices were identified by investigating the covariance structure (Row *et al.* 2017). For each of three cost surface and four genetic distance response variables, we fitted MLPE models that considered as fixed effects Euclidean distance only, LCP and Euclidean distance, and CT and Euclidean distance, resulting in a comparison of 36 MLPE models in total. We added Euclidean distance as a fixed effect to our LCP and CT MLPE models to account for the impact that isolation by geographic distance may have on the MLPE model outcomes (Row *et al.* 2017).

2.3.9 Functional landscape linkage maps in Circuitscape

Out of the 36 MLPE models, we identified a single model that had the lowest AIC, and the highest conditional, and marginal R^2. We independently compared AIC values per individual response variable since AIC values should not be compared across models that utilize different response variables. For each of our response variables, we identified the model with the lowest AIC. Our MLPE model assessment consistently identified CT distances calculated for the extreme nonlinear transformation (c=8) as the connectivity measure that best explained the observed gene glow (D_{PS}) across the landscape for all of the individual response variables. We therefore used CT distances calculated in Circuitscape (Shah and McRae 2008) to estimate, delineate and visualize areas in the landscape that represent functional landscape linkages. In Circuitscape, we used the extreme nonlinear transformation (c=8) of the HSM as our cost surface, and incorporated GD as resistors to current flow between pairs of connected nodes (McRae 2006; Shah and McRae 2008). Our final FLL map thus considers suitable habitats based on spatial data, and also gene flow as the inverse of genetic distance, to delineate areas in the landscape that may be important for maintaining or restoring population connectivity.

2.4 RESULTS

We genotyped 142 elephant fecal samples for 9 highly variable nuclear DNA microsatellite loci (Table A.1). Using GD calculated in AIS, we found no correlation between genetic and geographic distances (partial Mantel test $r = -0.035$, $P \geq 0.065$). To visualize gene flow across the landscape, we interpolated genetic distances to form a landscape shape in the program AIS. The interpolated GD surface is a relative representation of gene flow across the study area, since the GD estimates are relative to genetic variation captured by the microsatellite marker system. Our GD surface indicated that regions in South Africa (e.g., areas within Kruger

National Park) had higher gene flow compared to gene flow elsewhere in the study area (Figure 2.5).

2.4.1 Maxent model results

We found no significant correlation (Pearson's R <0.5, p >0.5) between our MaxEnt environmental layers when including "ambient population" presence rather than gradient-based distance from high human densities (Figure 2.6). We tested four regularization beta parameters in Maxent (β = 1, 2, 3, 4) for each of the three background datasets (entire extent, range-based, and use-based background). Analyses where background data were generated across environments that are not present in occurrence datasets (e.g., the entire extent background dataset in this study) can produce artificially inflated AUC values. In agreement, MAxEnt models that included background data from the entire extent had the highest AUC (Table 2.1), likely indicating inflated AUC values. Regardless of the AUC performance, background data should always be chosen to reflect the environmental conditions that are relevant to the species for which the model is generated. Comparing MaxEnt models that included range-based and use-based background datasets which reflect actual environmental conditions that the elephants encounter, we find that the range-based background dataset resulted in slightly higher AUC than use-based background dataset. We therefore selected the range-based dataset because it is less likely to be subject to artificial model inflation and outperformed the use-based dataset. We found that β = 1 was the regularization betaparameter that had the highest regularization gain and AUC, and we therefore selected as our HSM the MaxEnt model calculated using range-based background data and β = 1 (Figure 2.7).

We investigated how each of the environmental variables contributed to the predicted HSM and found that human density as "ambient population" presence, distance to water and

slope mostly contribute at low variable values (low human densities, close to water, and habitats with low slopes), and all three of these environmental variables decrease in their contribution until they reach a threshold environmental value beyond which they do not contribute to the predictive model (Figure A.4). The HSM indicates that elephants occur in areas of low human density (there is a negative relationship between human density and elephant presence), in areas that are close to water (there is a negative relationship between distance to water and elephant presence), and in areas with no or low slopes (there is a negative relationship with between slope and elephant presence). Intermediate values of primary productivity contributed the most to our predictive HSM model, which is congruent with the distribution of elephants in our study area that mostly occur in landscapes with intermediate productivity (e.g., savanna or non-woody habitats; Mapaure and Campbell, 2002; Young, Ferreira and Van Aarde, 2009).

2.4.2 Spatial and genetic connectivity across landscape resistance surfaces

2.4.2.1 MLPE models

We linearly and nonlinearly transformed the MaxEnt HSM (range-based background, $\beta =$ 1) into three alternative representations of landscape resistance. Our linear transformation (Figure 2.8A) assumes that resistance is inversely related to habitat suitability, while the nonlinear transformation by means of a negative exponential function allowed for the possibility of nonlinear responses that elephants might have to unsuitable habitats (Figure 2.8B & C).

For each of three cost surfaces and four genetic distance response variables, we fitted MLPE models that considered as fixed effects Euclidean distance only, LCP and Euclidean distance, and CT and Euclidean distance, resulting in 36 MLPE individual models but were only models that were based on the same response variable were compared. Based on AIC, conditional R^2 and marginal R^2, we find that CT and Euclidean distance calculated using the

pronounced nonlinear HSM transformation are the best predictors of D_{PS} as a proxy of gene flow across the landscape for all of our response variables (Table A.3).

2.4.2.2 Functional landscape linkage map

We used CT distances to estimate, delineate and visualize areas in the landscape that represent functional landscape linkages in Circuitscape. We used the pronounced nonlinear transformation of the HSM as our cost surface, and incorporated GD as resistors to current flow between pair of connected nodes to create a single FLL map that considered suitable habitats based on spatial data, and also gene flow as the inverse of genetic distance (Figure 2.9). This FLL map provided a more precise delineation of important linkages between landscape compared to connectivity maps produced using only spatial data (Figure 2.10 & Figure A.5). FLL may be important for delineating fine scale linkages across the landscape, where spatial data alone may not provide a high enough resolution for demarcating fine scale landscape linkages (Figure 2.10 & Figure A.5).

We identified four possible FLL that connected elephant populations across our study area. These FLL included linkages between Northern Botswana and Kafue National Park in Zambia (Figure 2.11 - oval 1 in the left panel), between Northern Botswana and Mapungubwe National Park in South Africa (Figure 2.11 - oval 2 in the left panel), between Mapungubwe National Park and north of Kruger National Park in South Africa (Figure 2.11 - oval 3 in the left panel), and between the Kruger National Park and elephant populations in Tembe (South Africa) and Maputo (Mozambique) Elephant Reserves (Figure 2.11 - oval 4 in the left panel). Our FLL map provides fine-scale delineation of landscape linkages. For example, moving from north to south, linkage 2 that spans the area between Northern Botswana and Mapungubwe National Park in South Africa initially consists of 3 routes that circumvent areas of high human presence and a

large water body (Sashe dam), but south of these barriers consists of a single route that follows the Botswana-Zimbabwe border closely until it reaches South Africa.

2.5 DISCUSSION

The defragmentation of conservation areas through the development and maintenance of FLL could induce regional demographic stability in elephant numbers, enhance seasonal changes in space use, reduce local impact that elephants have on the landscape (van Aarde and Jackson 2007), and could mitigate the consequences of genetic isolation by promoting gene glow and increasing genetic diversity (Allendorf and Luikart 2009; Burkart *et al.* 2016; Orton *et al.* 2020). Integrating multifaceted landscape habitat modeling with genetic analyses has been proposed and applied as a conservation tool, and is an expedient method through which FLL can be delineated. For example, the integration of spatial and genetic analyses has been used to detect and evaluate landscape connectivity and movement corridors for wolves (Kabir *et al.* 2017), Cantabrian brown bears (Mateo-Sánchez *et al.* 2015), rodents (Wang *et al.* 2008) and birds (Klinga *et al.* 2019). In this study, we show that the integration of spatial landscape modeling and genetic analyses can also be used to delineate FLL for African elephant conservation planning.

FLL rely on the accurate identification of environments that provide for the habitat preferences and requirements of the species of interest. Habitat preference is the ratio of use of a specific habitat over the availability of that habitat, and is conditional on the availability of all habitats present (Aarts *et al.* 2008). Maxent takes into account the availability of landscape variables by considering the frequency of variable occurrence (Merow *et al.* 2013). Because MaxEnt uses presence data only, the output predictions should be interpreted as a relative, rather than an absolute, indicator of habitat suitability (Elith *et al.* 2011). Our HSM predicted in MaxEnt accounts for functional responses to habitat suitability. Accounting for functional

responses is important since elephant habitat suitability has been shown to be dependent on habitat availability (Roever *et al.* 2012). Through the use of linear and nonlinear transformations of the HSM we were able to assess how important habitats of different suitability are for elephant space use and gene flow.

We find a pronounced nonlinear response to habitat suitability which suggests that elephant movement and gene flow are mostly impacted by very unsuitable habitats, and that moderately unsuitable habitat impedes connectivity between elephant populations to a lesser degree. This nonlinear response agrees with space use described for some elephant populations in our study region, for example, where elephants in Botswana are known to range across densely populated areas to access food and water (Hoare and Du Toit 1999; Jackson *et al.* 2008). The pronounced nonlinear response to occurrence-based habitat suitability models suggests that alternative quantifications of habitat use might be needed to accurately establish elephant habitat requirements. Researchers could, for example, separate elephant movements into different behavioral states using hidden Markov models fitted in a Bayesian framework (Leos-Barajas and Michelot 2018; Wang 2019; Vogel *et al.* 2020), and identify from those states the environmental variables that are important for establishing or maintaining connectivity. For example, Keeley *et al.*, 2017 show that habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements in kinkajous (*Potos flavus*), where tolerance for unsuitable habitat during dispersal seems common. Mateo-Sánchez *et al.*, 2015 show that dispersing Cantabrian brown bears might be more flexible in their dispersal movement behavior than they are in their habitat resource utilization behavior. Determining behavioral states may be especially relevant to elephant FLL use since Vogel *et al.*, 2020 show that within corridors, exploratory movements that are fast and directional are dominant compared to encamped foraging movements that are

slow and meandering. It may therefore be beneficial for future studies to consider both behavior and space use when delineating FLL for African elephants.

Our analyses indicated that landscape representations that consider multiple paths of connectivity (isolation by resistance based on circuit theory; IBR) perform better at predicting gene flow across the landscape than singular paths (LCP) or geographic distance alone. Similar IBR frameworks have been more effective at explaining gene flow for black bears (Cushman *et al.* 2006), hedgehogs (Braaker *et al.* 2017), and other plant and animal species (McRae and Beier 2007). Our FLL map shows that gene flow and connectivity appear to be influenced by different environmental variables across the landscape. We find areas in our FLL landscape where slope (Figure A.6), water availability (Figure A.7), "ambient human population" presence (Figure A.8) and primary productivity (Figure A.9) respectively seem to be the primary drivers of connectivity. When demarcating FLL for conservation, researchers and conservation stakeholders should consider that FLL and connectivity may be dependent on the specific geographic area and environmental variables under consideration, and that a "one-map-fits-all" approach should be avoided.

Our FLL map shows that analyses which incorporate genetic information when mapping connectivity could provide more precise delineations of linkages across the landscape. This is illustrated by comparing linkage maps that are based on only spatial data to linkage maps that incorporate both spatial and genetic data (Figure 2.10 & Figure A.5). For example, the landscape linkage between Northern Botswana and Mapungubwe National Park in South Africa does not contain clearly delineated routes when considering only spatial data, but adding gene flow to the connectivity analyses resulted in fine-scale route delineations (Figure 2.11).

The largest landscape linkage identified in this study extends from Northern Botswana to Mapungubwe National Park in South Africa (Figure 2.11). Moving from north to south, this linkage initially consists of 3 routes that circumvent areas of high human presence and a large water body (Sashe dam), but south of these barriers consists of a single route that follows the Botswana-Zimbabwe border closely until it reaches South Africa. When demarcating areas for elephant conservation in this region, conservation stakeholders may consider whether there are alternative routes that link the areas of interest (e.g. the three routes in the northern portion of linkage 2), and prioritize routes and linkages where landscape connectivity seems most vulnerable to change.

Researchers and conservation stakeholders should also consider factors other than landscape connectivity and gene flow when demarcating landscape linkages for conservation. For example, factors such as FLL land ownership (Pinter-Wollman 2012), human-elephant conflict mitigation (Jackson *et al.* 2008; Pinter-Wollman 2012), the direct and indirect impact of FLL creation on local indigenous communities (Baldus *et al.* 2007), FLL overlap with poaching hotspots (Osborn and Parker 2003; Booth and Dunham 2016), and other sociopolitical factors need to be integrated into conservation decisions.

2.5.1 Future directions

Elephant space use and habitat suitability modeling may be influenced by factors that were not considered in this study. For example, Mashintonio *et al.* (2014) show that elephants select habitat based on environmental qualities at multiple scales, and it may therefore be informative to incorporate habitat suitability modelling approaches other than MaxEnt that could allow for the consideration of environmental variables at different scales. In addition, the spatial resolution of predicted HSM can influence mixed effect models and researchers may therefore

want to compare predictor outcomes for different spatial resolutions by upscaling and downscaling resistance surfaces for each of the HSMs before evaluating them with connectivity metrics. Researchers could, for example, use an Gaussian pixel smoothing algorithm approach which can be effective for determining the scale at which elephants select resources (Mashintonio *et al.* 2014). In addition to scale, seasonality and sex are crucial drivers of elephant dispersal and land use patterns (Young *et al.* 2009a). Future studies should therefore consider seasonal (temporal) and sex-specific differences when predicting gene flow across the landscape, where individually modelled FLL connectivity maps for males and females, and for wet and dry seasons could be generated. Despite these limitations, the novel conservation genetic approach developed and applied here integrates multidisciplinary data and methods in an innovative way to inform conservation planning, and may therefore still be useful as a framework for future studies on elephant and other taxa that aim to develop spatially and genetically informed conservation strategies.

2.6 TABLE AND FIGURES

2.6.1 Table

Table 2.1 MaxEnt models that included background data from the entire geographic extent of our study area had the highest area under the receiver-operator curve (AUC), while the range-based background dataset resulted in slightly higher AUC than use-based background dataset.

Background data	Maxent regularization parameter	Regularized training gain	AUC
Entire extent	$\beta = 1$	0.060	0.8568687
	$\beta = 2$	0.052	0.854843
	$\beta = 3$	0.046	0.8524966
	$\beta = 4$	0.040	0.8492671
Range-based background data[1]	$\beta = 1$	0.039	0.7860084
	$\beta = 2$	0.034	0.7808212
	$\beta = 3$	0.03	0.7745906
	$\beta = 4$	0.027	0.7644965
Use-based background data	$\beta = 1$	0.039	0.7583397
	$\beta = 2$	0.035	0.7544933
	$\beta = 3$	0.032	0.7505779
	$\beta = 4$	0.030	0.7448165

[1] We selected the MaxEnt model that was calculated using the range-based background dataset and a regularization parameter of $\beta = 1$ for subsequent analyses because it was less likely to be subject to artificial model inflation associated with the entire extent background dataset, and because it outperformed the use-based dataset.

2.6.2 Figures

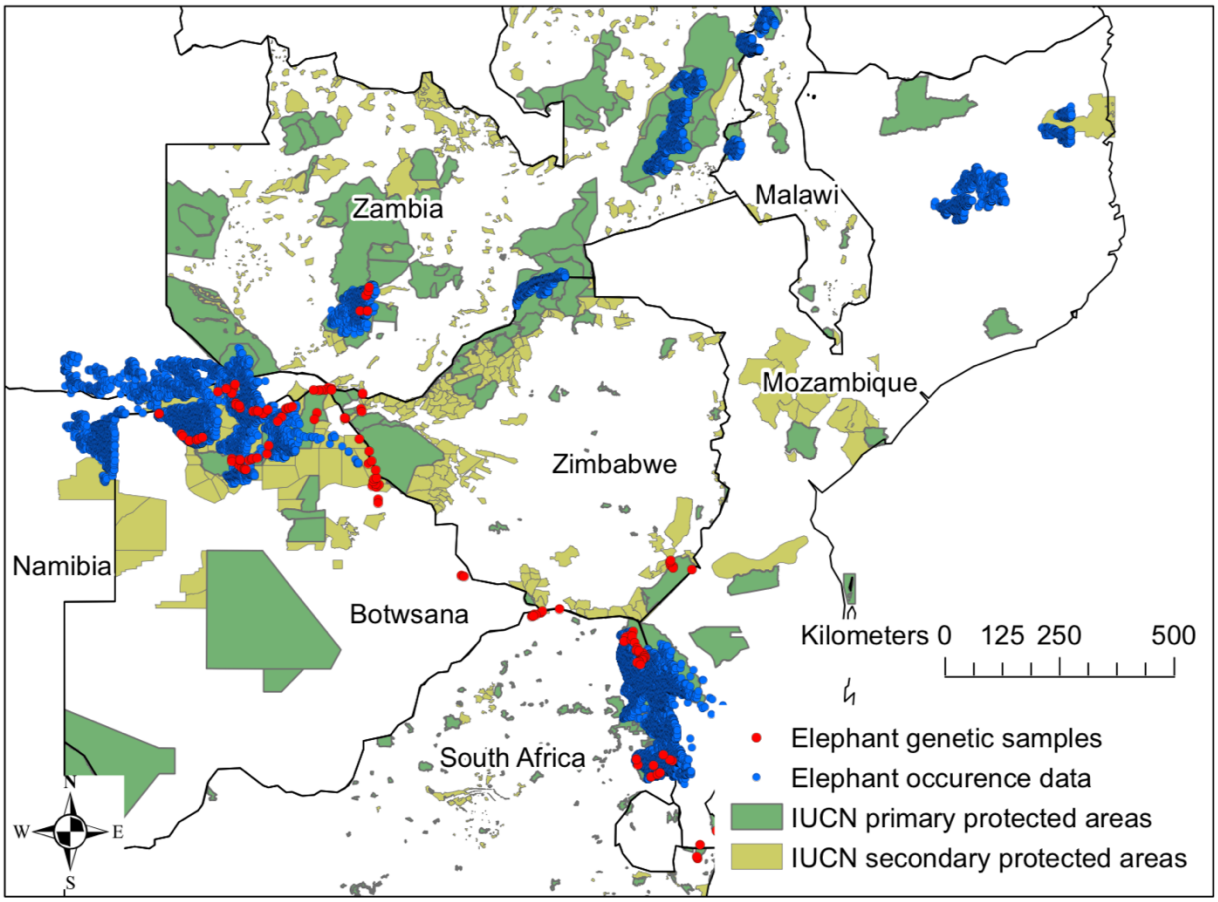

Figure 2.1 Our study area spans seven countries across southern Africa, including Botswana, Malawi, Mozambique, Namibia, South Africa, Zambia and Zimbabwe that collectively contain >70% of the number of elephants in Africa and >42% of the total range of elephants in Africa (Thouless *et al.* 2016). Occurrence data (GPS telemetry points) from 116 elephants are shown in blue, and locations at which genetic samples were collected are shown in red.

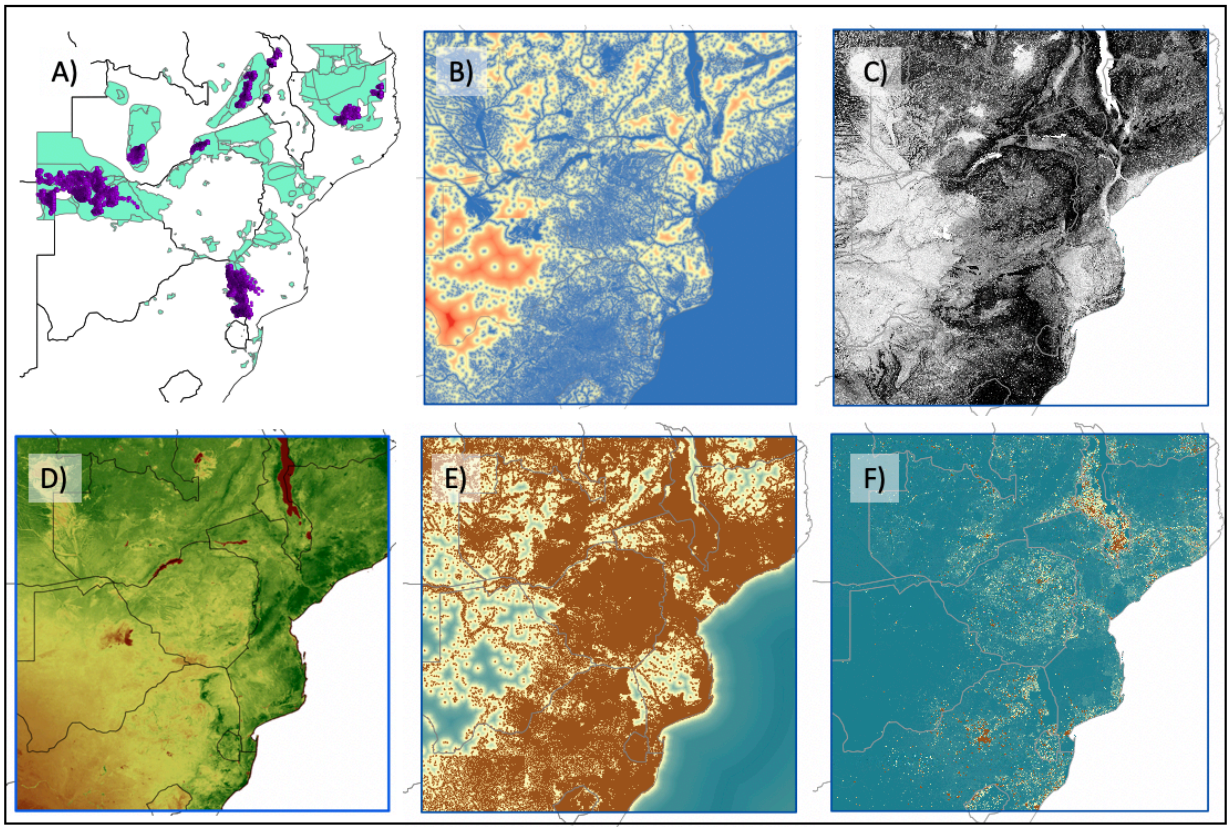

Figure 2.2 Our MaxEnt habitat suitability model (HSM) was based on elephant occurrence points that correspond with areas of known elephant range (panel A - purple dots represent occurrence data and green areas represent the current elephant range as demarcated by the IUCN Red List of Threatened Species; Blanc, 2008). MaxEnt environmental variables included habitat covariates that are known to influence elephant space use: water availability as represented by distance to the closest water source (panel B – blue areas are close to water, red areas are far from water), the slope of the land (panel C – dark areas represent steep slopes), primary productivity indexed by long-term mean Enhanced Vegetation Index (EVI; panel D – green areas indicate high primary productivity, brown and purple areas indicate low primary productivity), and human presence as a transformed gradient-based metric (panel E – red indicates areas close to high human densities, blue indicates areas that are far away from high human densities) and as an ambient population presence (panel F – red indicates areas with high ambient population presences, blue indicates areas with low ambient population presences).

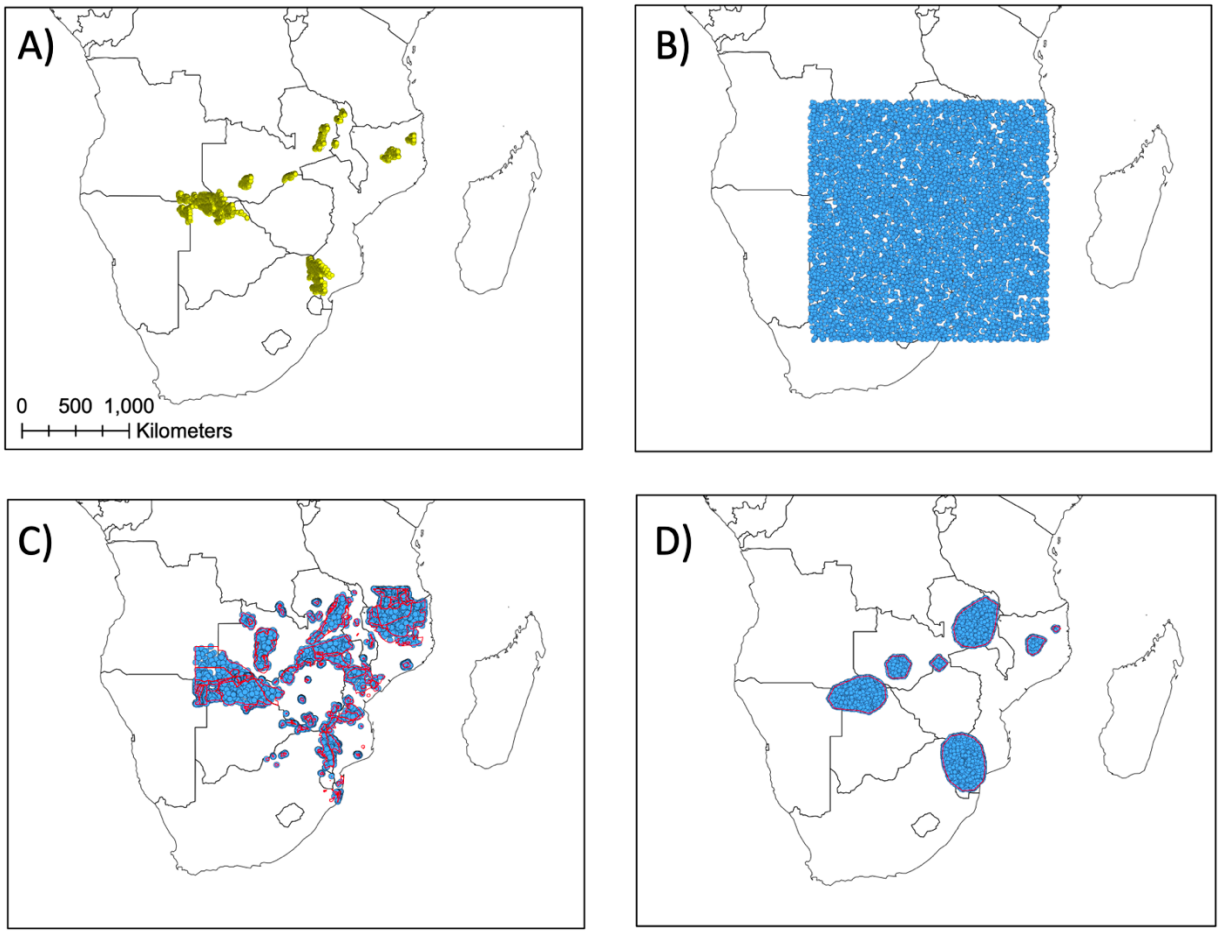

Figure 2.3 MaxEnt compares environmental conditions at background location data points to environmental conditions at known presence (occurrence) locations (panel A – yellow dots indicate occurrence points). Background data should reflect the environmental conditions that are relevant to the species for which the model is generated. We tested three different background datasets: background data generated for the entire extent of all environmental predictors included in the model (panel B); range-based background data generated for areas of known elephant range as demarcated by the IUCN Red List of Threatened Species (panel C – red lines indicate known range; Blanc, 2008); and use-based background data from areas included in an 80% Convex Hull (CH) of presence points (panel D – red lines indicate the 80% CH). For each of these datasets we generated 10 000 random points (shown in blue in panels B, C and D) to compare to occurrence data.

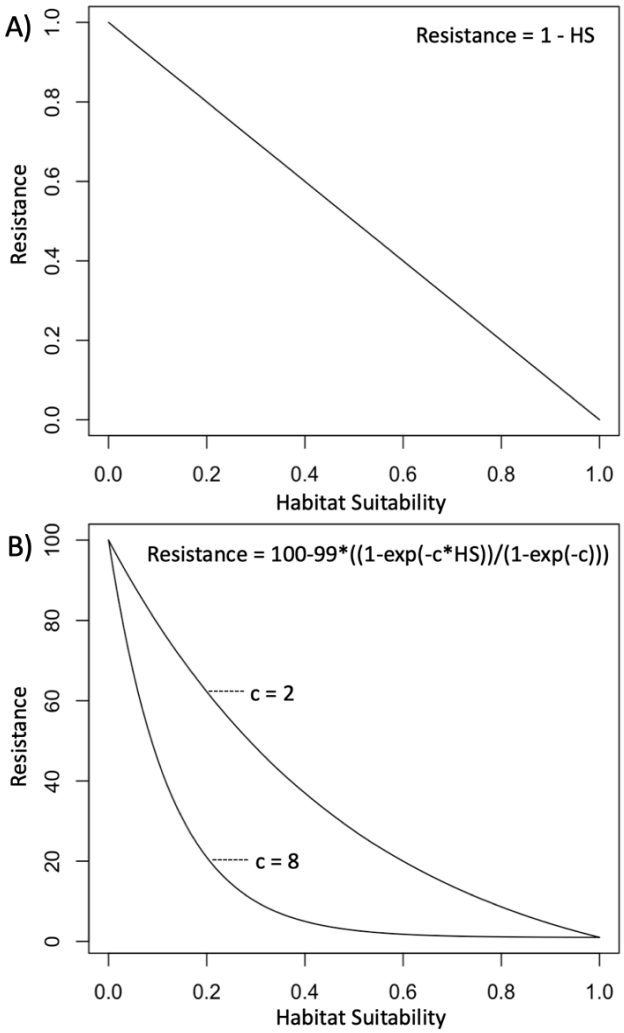

Figure 2.4 The single best habitat suitability model (HSM) was transformed linearly (panel A) and nonlinearly (panel B) into resistance surfaces. A linear transformation assumes that resistance is inversely related to habitat suitability, where less suitable habitat would be costlier to move through. A transformation by means of a negative exponential function allowed for the possibility of nonlinear responses that elephants might have to unsuitable habitats. We generated resistance surfaces for transformations using c = 2 and c = 8 to respectively generate slight and pronounced nonlinear responses. Equations for linear and nonlinear transformations are provided in the upper right corner of each graph.

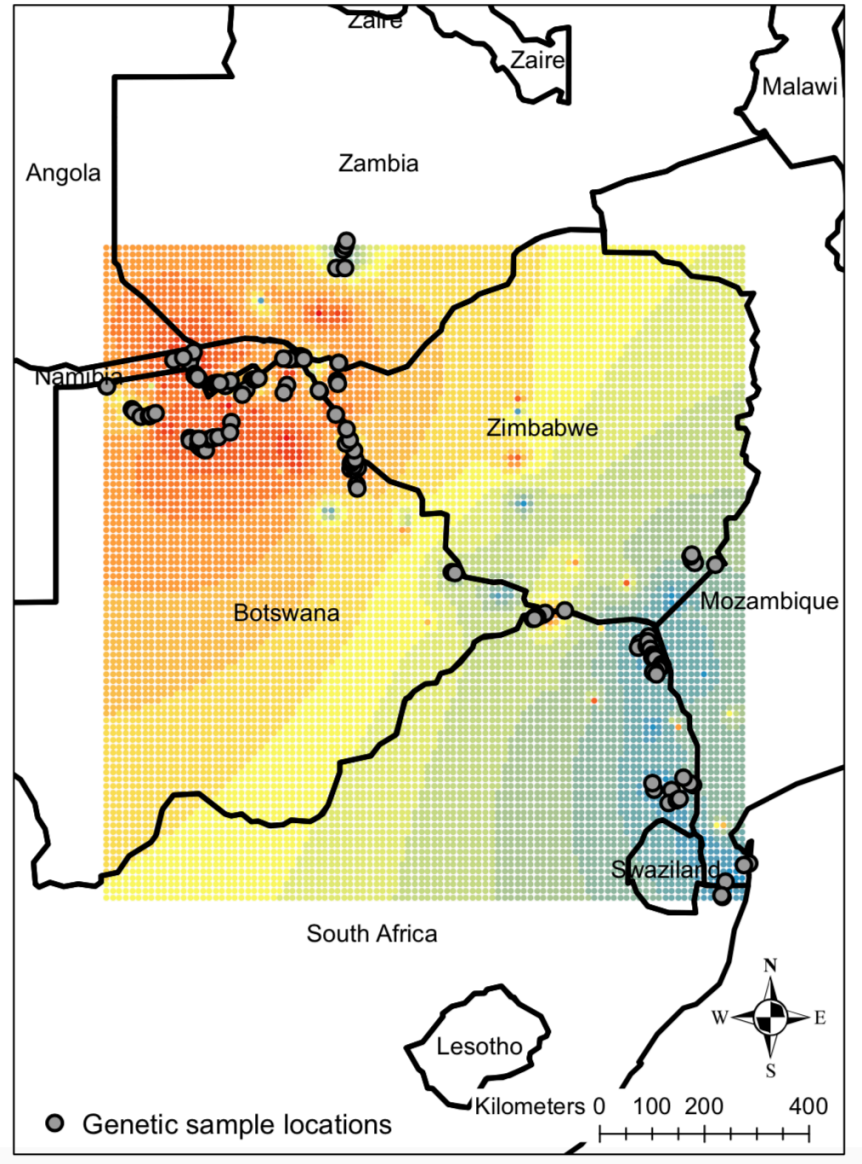

Figure 2.5 Our visual representation of gene flow across the study extent indicated that areas in South Africa (e.g. Kruger National Park) had higher gene flow (low interpolated genetic distances) relative to gene flow elsewhere. The interpolated genetic distance surface is a relative representation of gene flow across the study area and is dependent on the genetic variation captured by the marker system. Areas in blue indicate higher relative gene flow (lower genetic distance) and areas in red indicate lower relative gene flow (higher genetic distance) among 142 elephant genetic samples (grey circles).

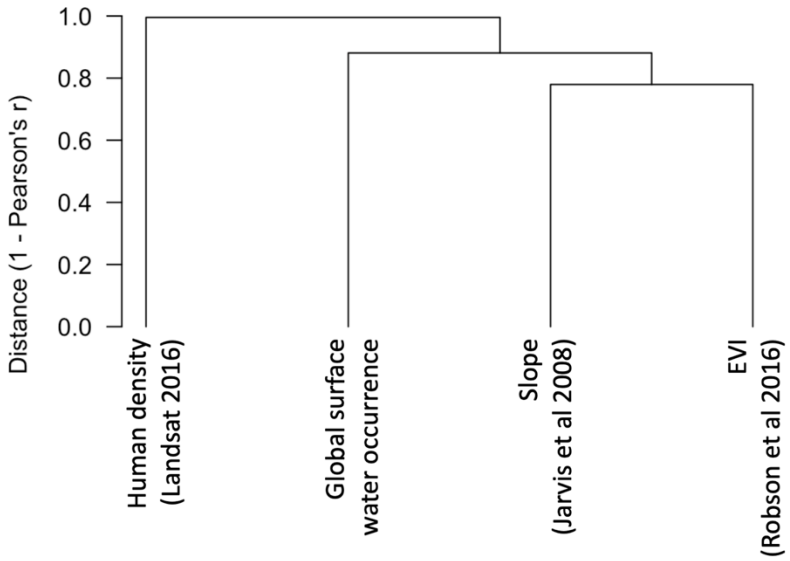

Figure 2.6 We found no significant correlation between any of the environmental variables included in the MaxEnt model when using "ambient population" presence as our human density indicator. We tested Pearson's R cut-off values of 0.7 (which is the least stringent and allows for most collinearity), 0.6, and 0.5 (the most stringent, presented here) and found no significant collinearity for any of the values.

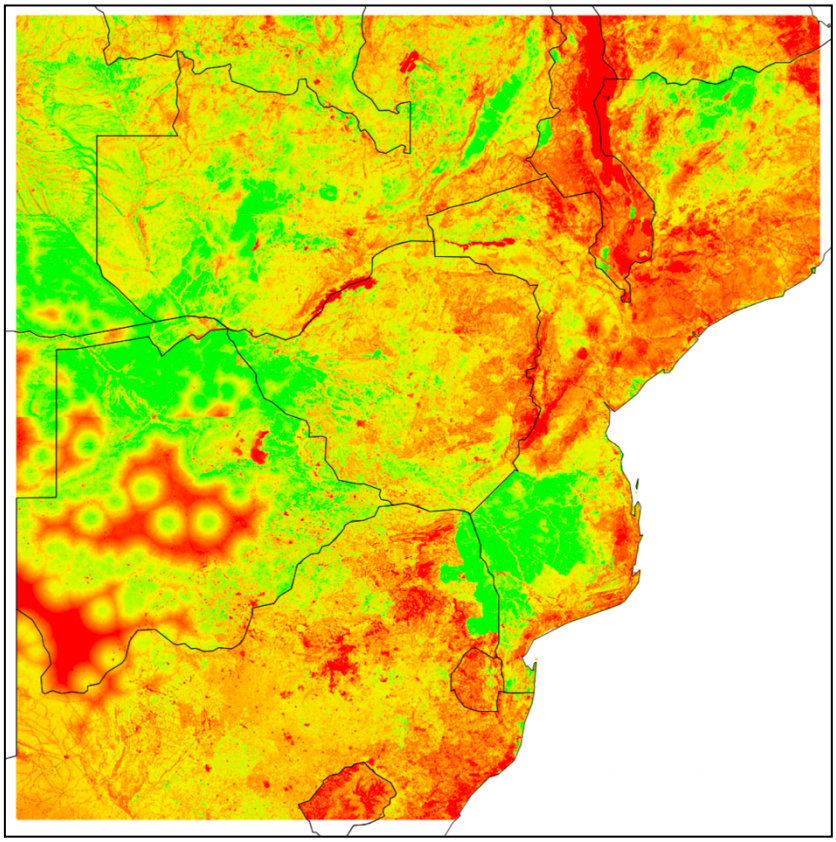

Figure 2.7 The habitat suitability model (HSM) was predicted in MaxEnt using range-based background data and a regularization betaparameter of 1. Suitable elephant habitat is shown in green while less suitable habitats are shown in shades of yellow and red. This HSM was subsequently transformed linearly and nonlinearly into resistance surfaces.

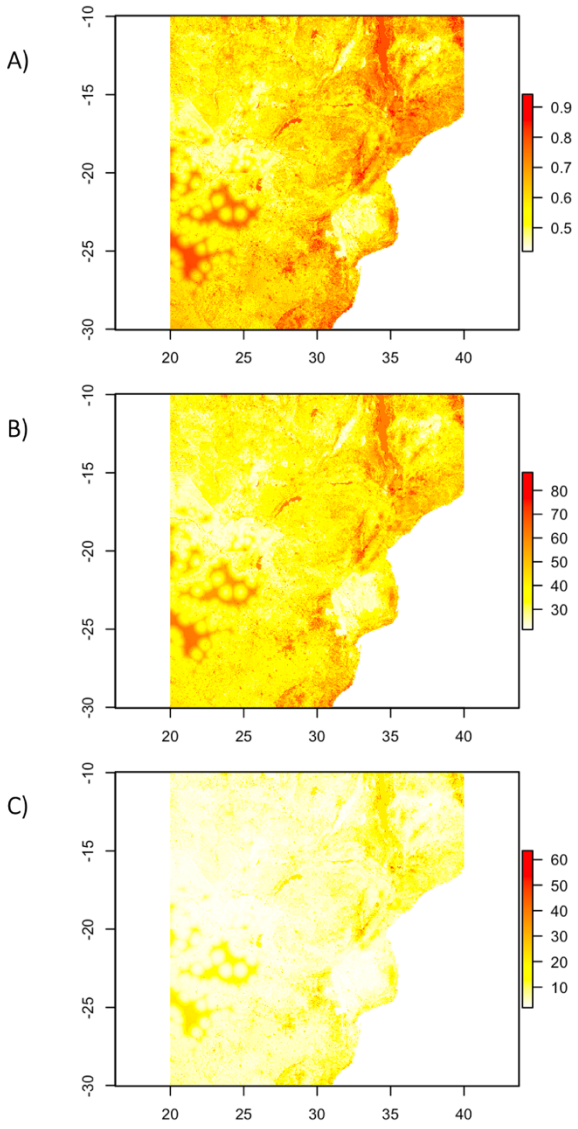

Figure 2.8 We transformed the MaxEnt habitat suitability model (HSM) into three alternative representations of landscape resistance. Our linear transformation (Figure 2.8A) assumes that resistance is inversely related to habitat suitability, while the nonlinear transformations allowed for slight nonlinear responses (Figure 2.8B) and pronounced nonlinear responses (Figure 2.8C) to unsuitable habitats. The X and Y axes indicate longitude and latitude, and map colors indicate areas in the landscape that are very costly (red) to less costly (white) for elephants to move through.

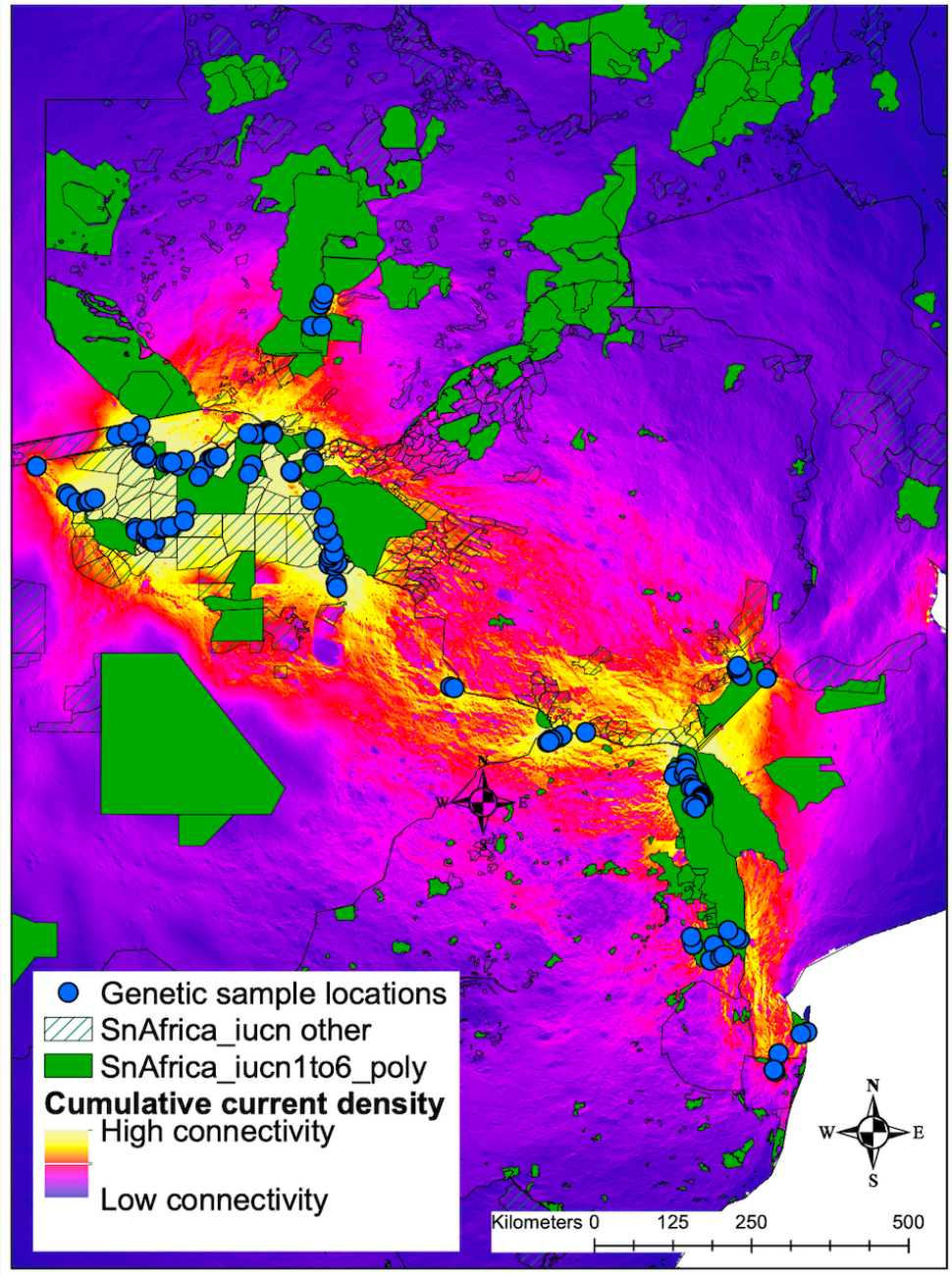

Figure 2.9 We used the software Circuitscape to estimate, delineate and visualize areas in the landscape that represent functional landscape linkages (FLL). Our cumulative current density map of FLL considers suitable habitats based on spatial data, and also gene flow as the inverse of genetic distance, to delineate areas in the landscape (yellow) that may be important for maintaining or restoring population connectivity. Green areas indicate IUCN primary protected areas, cross hatched areas indicate IUCN secondary protected areas and blue dots indicate the genetic sample locations among which connectivity was modelled.

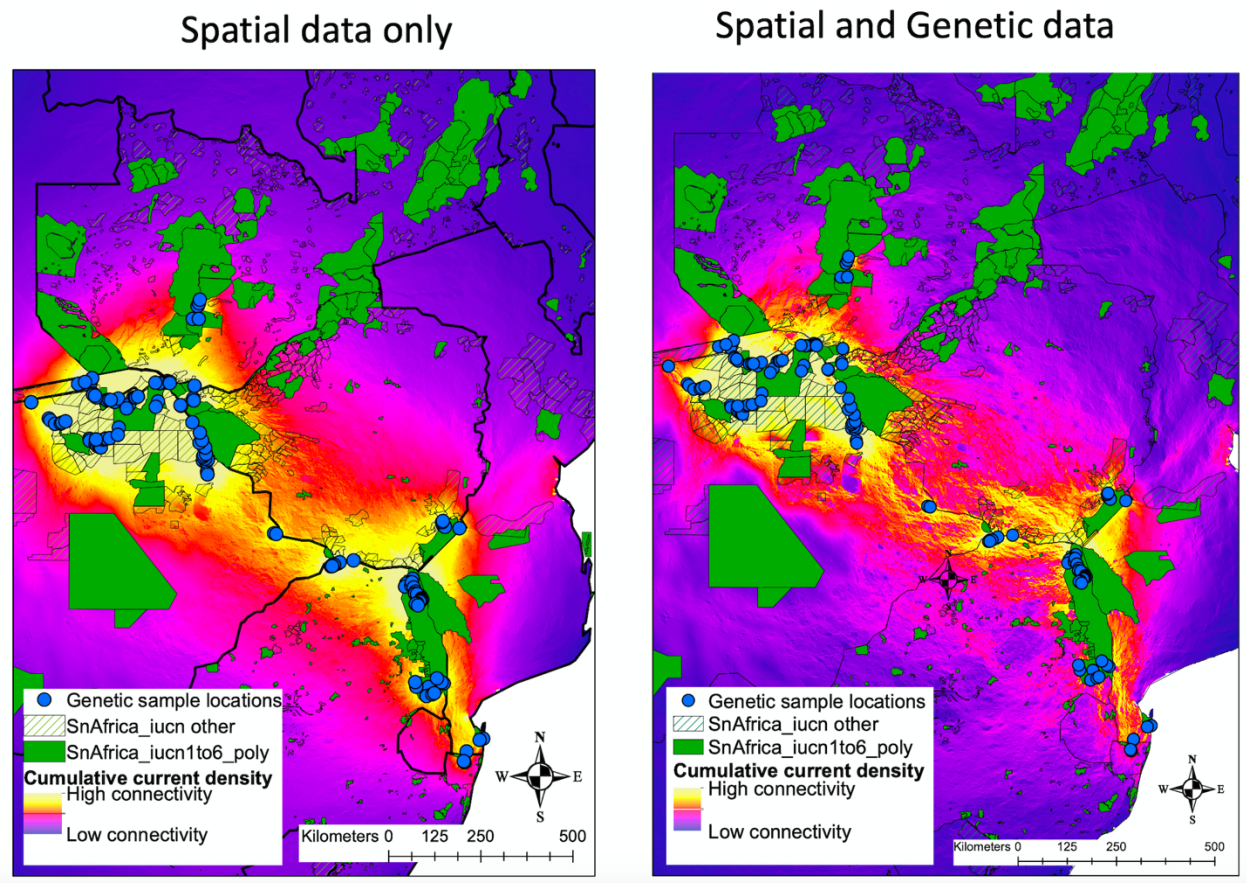

Figure 2.10 FLL maps that are based on both spatial and genetic data (right panel) provide more precise delineations of landscape linkages compared to connectivity maps produced using only spatial data (left panel).

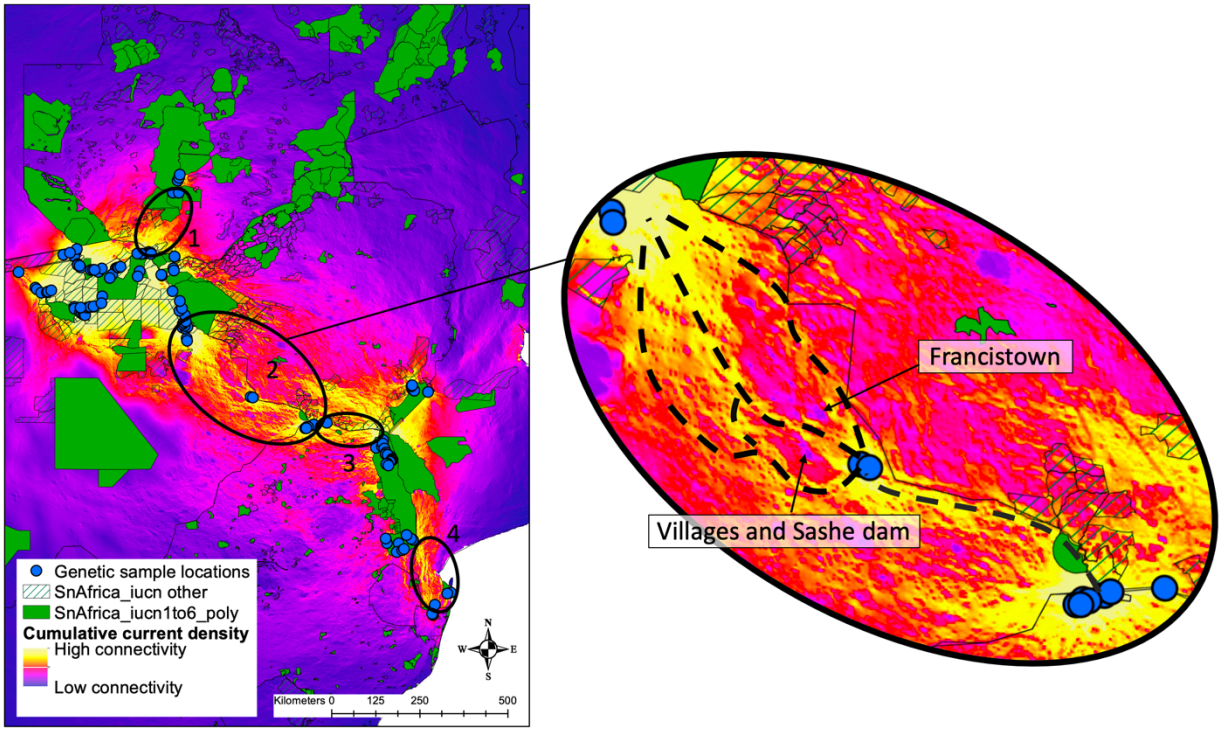

Figure 2.11 We identified four possible FLL that connected elephant populations across our study area. These FLL included landscape linkages between Northern Botswana and Kafue National Park in Zambia (oval 1 in the left panel), between Northern Botswana and Mapungubwe National Park in South Africa (oval 2 in the left panel), between Mapungubwe National Park and north of Kruger National Park in South Africa (oval 3 in the left panel), and between the Kruger National Park and elephant populations in Tembe (South Africa) and Maputo (Mozambique) Elephant Reserves (oval 4 in the left panel). Our FLL map provides fine-scale delineation of landscape linkages. For example, moving from north to south, linkage 2 that spans the area between Northern Botswana and Mapungubwe National Park in South Africa initially consists of 3 routes that circumvent areas of high human presence and a large water body (Sashe dam), but south of these barriers consists of a single route that follows the Botswana-Zimbabwe border closely until it reaches South Africa. Green areas indicate IUCN primary protected areas, cross hatched areas indicate IUCN secondary protected areas and blue dots indicate the genetic sample locations among which connectivity was modelled.

CHAPTER 3: ORIGIN AND PHYLOGEOGRAPHY OF AFRICAN SAVANNAH ELEPHANTS (*LOXODONTA AFRICANA*) IN KRUGER AND NEARBY PARKS IN SOUTHERN AFRICA[1]

3.1 ABSTRACT

African savannah elephants (*Loxodonta africana*) occur in fragmented and isolated populations across southern Africa. Transfrontier conservation efforts aim at preventing the negative effects of population fragmentation by maintaining and restoring linkages between protected areas. We sought to identify genetic linkages by comparing the elephants in Kruger National Park (South Africa) to populations in nearby countries (Botswana, Mozambique, Zambia and Zimbabwe). We used a 446 base pair mitochondrial DNA (mtDNA) control region fragment (141 individuals) and 9 nuclear DNA (nDNA) microsatellite markers (69 individuals) to investigate phylogenetic relationships and gene flow among elephant populations. The mtDNA and nDNA phylogeographic patterns were incongruent, with mtDNA patterns likely reflecting the effects of ancient female migrations, with patterns persisting due to female philopatry, and nDNA patterns likely reflecting male-mediated dispersal. Kruger elephant heterozygosity and differentiation were examined, and were not consistent with genetic isolation, a depleted gene pool or a strong founder effect. Mitochondrial DNA geographic patterns suggested that the Kruger population was founded by elephants from areas both north and south of Kruger, or has been augmented through migration from more than one geographic source. We discuss our findings in light of the need for conservation initiatives that aim at maintaining or restoring connectivity among populations. Such initiatives may provide a sustainable, self-

[1] Chapter 3 has previously been published as de Flamingh, A., Roca, A. L., & Van Aarde, R. J. (2018). Origin and phylogeography of African savannah elephants (Loxodonta africana) in Kruger and nearby parks in southern Africa. *Conservation Genetics*, *19*(1), 155-167. It is reproduced here in adherence to copyright guidelines.

regulating management approach for elephants in southern Africa while maintaining genetic diversity within and gene flow between Kruger Park and nearby regions

3.2 INTRODUCTION

Landscape fragmentation and habitat loss are globally important drivers of population declines among species (Lindenmayer and Fischer 2007). Protected areas may serve as species strongholds, but they often consist of discontinuous patches that represent fragments of the once continuous historic ranges of many mega-herbivores and carnivores (Ripple *et al.* 2015). The fragmentation and resulting isolation of populations may have deleterious demographic and genetic consequences, because restricted gene flow between populations may lead to genetic drift, inbreeding, the expression of deleterious alleles, and reduced fitness; it can also increase the risk of local extirpation (Allendorf *et al.* 2013). For example, Fitzpatrick and Evans, 2009 showed that reduced heterozygosity impaired sperm quality in endangered mammals, and (Miller *et al.* 2011) showed that island tammar wallaby populations with low genetic diversity and high levels of inbreeding had an increased frequency of morphological abnormalities.

Recent conservation initiatives in Africa have sought to counter the effects of population fragmentation and isolation. These initiatives rely on the restoration and maintenance of distributional ranges to link formerly isolated populations through the creation of "megaparks" (van Aarde and Jackson 2007), and transfrontier conservation areas (Hanks 2003). Such initiatives can be considered functional if they disrupt population isolation by allowing dispersal or migration between nearby populations. Linking isolated populations may have beneficial genetic consequences if isolation has been a driving force of genetic degradation associated with bottlenecks and the subsequent loss of genetic diversity due to drift.

Information on the genetic structure and characteristics of elephant populations may help establish whether isolation and founder effects have resulted in genetically depleted, homogenous populations, and may provide support to the proposed benefits of enabling dispersal between isolated populations. African savannah elephants are no longer distributed across a continuous geographic range, but instead have contracted ranges that largely overlap with protected areas (Ripple *et al.* 2015). Elephants in Kruger National Park in South Africa are confined by fencing, while in parts of Angola, Botswana, Namibia, Mozambique and Zimbabwe elephant movements are restricted by the abundance of people or by resource limitations (Roever *et al.* 2013). Gene flow between Kruger and nearby parks has not been quantified, but is thought to be limited due to park fences and other barriers (e.g., areas of high human density) that obstruct dispersal between populations. Furthermore, the Kruger elephant population is believed to have derived from a single source population. Elephants were widely distributed across southern Africa, including the present-day location of Kruger, and their movements were largely unrestricted prior to European colonization (Whyte 2001). After colonization hunting decimated the population in Kruger to the point of extirpation or near extirpation (Pienaar 1963; Hall-Martin 1992; Whyte 2001). It is believed that the population was re-established by elephants that migrated from Mozambique into Kruger at an entry point approximately midway between the Park's northern and southern boundaries after the proclamation of the Sabie Game Reserve in 1898 and the arrival of the first warden in 1903. Elephants are believed to have migrated from this entry point into the northern and southern regions of the park (Whyte 2001). The influx of elephants from Mozambique increased the population in Kruger to about 25 individuals in 1912 and about 100 individuals in 1926. By 1963 the population was estimated to include 1750

individuals and by the end of the 20th century the number of elephants in Kruger was fast approaching the 10 000 mark (van Aarde *et al.* 1999; Young *et al.* 2009b).

Our present study examined the genetic diversity and connectivity of elephants in South Africa's Kruger National Park, comparing them to elephants in nearby parks in Botswana, Mozambique, Zambia and Zimbabwe. We amplified a 446 base pair (bp) mitochondrial DNA (mtDNA) control region fragment from 141 elephants, and 9 nuclear DNA (nDNA) microsatellite loci from 69 elephants.

The elephant dung samples were collected across various national parks and conservation areas in southern Africa (Figure 3.1). Translocations were used to limit the increasing elephant population in Kruger from 1970 to 1994 (Whyte 2001; Anthony and Avery 2008), and translocations from Kruger to other areas continue to the present day. However, translocations have mainly been to areas (parks, conservation areas and privately-owned land) in South Africa that did not form part of this study (Dublin 2003; Anthony and Avery 2008).

We expected to detect in Kruger the negative genetic consequences that may follow a founder effect and population fragmentation and isolation. Compared with other southern African elephant populations, we expected Kruger's elephant population to show reduced genetic variability, and to find evidence of limited gene flow between Kruger and nearby parks, and the signatures of a single founding population.

3.3 METHODS AND MATERIALS

3.3.1 Sample collection

The collection of samples of elephant dung was sanctioned by appropriate authorities prior to collection (see Table B.1 for a list of permits). Sampling locality abbreviations include SKNP and NKNP = southern and northern Kruger National Park, respectively, South Africa;

MAP, LMAP and BZIM = Mapungubwe National Park (MAP) and along the Limpopo River that connects Mapungubwe and Kruger National Park (LMAP and BZIM), South Africa; TEP and BMAP = Tembe Elephant Park, South Africa and Maputo Elephant Reserve, Mozambique; KF = collected from Kafue National Park, Zambia; LV = Livingstone Town, Zambia; PT = Pandamatenga, Botswana; MR = Moremi Game Reserve, Botswana; SS and SN = south and north Savuti Game Reserve, respectively, Botswana; CH = Chobe National Park, Botswana; SR = Seronga, Botswana; LY = Linyanti River, Botswana; BZ = the border of Botswana and Zimbabwe; HW = Hwange National Park, Zimbabwe; GR = Gonarezhou National Park, Zimbabwe; and CP = Caprivi region, Namibia.

For a comparison of nDNA heterozygosity by geographic region, the samples were grouped geographically into 7 regions including Chobe Region (CR) = CH, MR, CP, SS, SN, LY; Hwange Region (HR) = HW, PT; Livingstone Region (LR) = LV; Botswana-Zimbabwe border Region (BZR) = BZ; Gonarezhou National Park (GNP) = GR; Mapungubwe Region (MR) = MAP, LMAP, BZIM; Kruger Region (KR) = NKNP, SKNP. Regional heterozygosity was not calculated for samples from TEP, BMAP, KF, SR and LV since samples from those areas were not genotyped for nDNA. Samples were grouped to represent contiguous geographic areas: samples that were from contiguous national parks or reserves, e.g. Chobe, Savuti, Linyanti and Moremi were grouped together and assigned the name of the largest national park; samples that were in close proximity to a national park but fell outside its borders were grouped as part of that park (e.g. the grouping of LMAP and BZIM as part of MAP); and samples that were collected along the Botswana-Zimbabwe border were grouped together.

3.3.2 Sample grouping

Samples were grouped and analyzed as follows: 1) groups determined by Bayesian clustering analysis in Geneland (Guillot *et al.* 2005), 2) seven regions based on geography and connectivity for nDNA heterozygosity comparisons, namely CR, HR, LR, BZ, GR, MR, KR and 3) four regions (all samples north of Kruger, samples in northern Kruger, samples in southern Kruger, and all samples south of Kruger) to compare mtDNA haplotype distribution analysis between north Kruger, south Kruger and nearby parks.

3.3.3 Data generation

Dung sample collection, DNA extraction and DNA amplification procedures for mtDNA sequences and nDNA microsatellites are described by de Flamingh *et al.* (2015). However, microsatellite genotypes determined for this study were replicated 4 times for homozygotes and 3 times for heterozygotes. Mitochondrial DNA sequence data (446 base pairs) for 141 individuals and nDNA data for 69 individuals (Figure 3.1, Table B.1) were generated at the Conservation Ecology Research Unit (CERU), University of Pretoria.

The difference in sample size for mtDNA and nDNA data can be attributed to the use of degraded fecal DNA - see de Flamingh, Sole and van Aarde, 2015 for a description of sample collection procedures. Mitochondrial DNA is often better preserved than nDNA in degraded DNA samples (Schwarz *et al.* 2009). Therefore, DNA extraction yield for mtDNA is often much higher than nDNA, leading to a higher amplification success of mtDNA markers compared to nDNA markers. This is especially apparent in degraded DNA samples (Andréasson *et al.* 2002). In addition, microsatellite genotype errors are often associated with degraded DNA amplification, e.g. null allele or false allele amplification (Taberlet *et al.* 1999; de Flamingh *et al.* 2014). Combined these factors resulted in fewer individuals being genotyped for nDNA than

mtDNA. However, the nDNA dataset sample size (69 individuals) was larger than sample sizes used by other studies to represent the same geographic area; e.g., Comstock *et al.*, 2002 used 19 individuals, Eggert, Rasner and Woodruff, 2002 used 54 individuals, and Roca *et al.*, 2001 used 40 samples to represent the same geographic region.

Outgroup taxa for mtDNA phylogenetic analyses included sequences from the Asian elephant (*Elephas maximus*) (Rogaev *et al.* 2006; Maikaew *et al.* 2011) and the woolly mammoth (*Mammuthus primigenius*) (Krause *et al.* 2006; Gilbert *et al.* 2008; Enk *et al.* 2011). GenBank accession numbers for outgroup taxa are listed in Table B.2.

3.3.4 Phylogenetic analysis

Sequence data were aligned using the program MUSCLE (Edgar 2004). We visually inspected the aligned sequences, excluded all oligonucleotide primer sequences, and verified each polymorphic site by viewing chromatograms in the program CLC Bio Genomics workbench (CLC Bio, Cambridge, MA).

The substitution model that best describes the variation in our data was identified using the program jModelTest2 (Darriba *et al.* 2012). Based on the lowest likelihood scores for models that fall within the 95% confidence interval for the Akaike and Bayesian Information Criterion, we used the HKY+I+G substitution model because it was consistently ranked highly by both criteria. Genescan fragments for nDNA microsatellite loci were viewed and scored using GeneMapper â Software Version 3.7 (Applied Biosystems, Foster City, USA). Scored data were then exported in table format to Excel spreadsheets for further analysis.

Distinct haplotypes present among the 147 individuals (141 African elephants and 6 individuals from outgroup taxa) were identified using DnaSP 5.10.01 (Librado and Rozas 2009), and used to construct a median-joining network using the software Network 4.6.1.0 (Fluxux

Technology Limited). Haplotypes were compared to previously published control region sequences and assigned to one of eight mtDNA subclades using the methods described by Ishida *et al.* (2013). Haplotypes assigned to the same subclade were verified as grouping together in the mtDNA median-joining network. Haplotypes 1 and 15, which were separated from other haplotypes by long branches, were queried using BlastN to find matches to the sequences of Ishida *et al.* (2013)

To establish the origins of Kruger elephants, mitochondrial DNA haplotype frequencies were compared across elephants in parks north of Kruger, northern Kruger, southern Kruger, and parks south of Kruger. For this comparison the haplotypes were assigned to four regions: all parks north of Kruger, northern Kruger, southern Kruger and all parks south of Kruger. The proportion of elephants in each region that carried each distinct haplotype was visualized by a stacked bar graph produced in Excel 15.12.3 (©Microsoft) that showed the most common haplotypes at the top of each bar, with rare haplotypes at the base of the bar.

Bayesian analyses were performed using the program MrBayes 3.2.6 (Ronquist *et al.* 2012). We performed three independent runs. The first run included 1 million generations with a sampling frequency of 10 000, the second run included 1 million generations a sampling frequency of 100, and our third run included 1 000 generation sampling all possible trees. The latter two runs were used to investigate convergence and stationarity of log likelihood values. For both runs the parameters converged within the first ca. 1 000 generations, confirming that convergence and stationarity had been reached within the initial run. Subsequent Bayesian analyses included a generation time of 1 million and a sample frequency of 10 000. Runs incorporated four chains and the HKY+I+G substitution model. Run length sufficiency was determined using the program Tracer 1.6.0 (Rambaut *et al.* 2014) by confirming adequate chain

mixture and convergence and stationarity upon a likelihood score. The ESS value indicated that a burn-in of 10% was sufficient (convergence was reached quickly), and we therefore discarded 10% of the initial trees. As an alternative visualization of topology support, Bayesian output trees were plotted as a cloudogram using the program DensiTree 2.2 (Bouckaert and Heled 2014).

3.3.5 Phylogeographic analysis

We used Geneland (Guillot *et al.* 2005) to determine how many sub-populations (K) of genetically similar individuals were present in our study, for both the mtDNA and nDNA datasets. We tested the possibility of K = 1 through K = 10, using 500 000 iterations with a thinning factor of 100. We employed the correlated allele frequency and spatial models to map probabilities of population membership across the landscape, using a spatial uncertainty parameter of 0.01 that was calculated for the same data by de Flamingh, Sole and van Aarde (2015). We verified Geneland clustering results with the program BAPS (Corander *et al.* 2003) using the "spatial clustering of individuals" algorithm for mtDNA sequence data (Kmax 15) and nDNA microsatellites (Kmax range of 5, 10 and 15 with each value replicated 5 times).

Geneland uses geo-referenced multi-locus data, we therefore converted the sequence data to haploid data based on binary code, where all the polymorphic sites (n=47) were considered as individual loci, with the possibility of 4 alleles (ACTG) at each of the loci (Guillot *et al.* 2005). We used an analysis of molecular variance (AMOVA) to calculate inter-population genetic variation (Fst – fixation index) for both nDNA and mtDNA. The AMOVA and exact test of population differentiation for mtDNA were implemented in Arlequin (Excoffier and Lischer 2010), and the AMOVA for nDNA was implemented using GenAlEx (Peakall and Smouse 2006). Observed nDNA heterozygosity was calculated for the seven regions (CNP, HNP, LNP, BZB, GNP, MNP, and KR) defined above using the program GenAlEx (Peakall and Smouse

2006). We used an AMOVA in GenAlEx (Peakall and Smouse 2006) and exact tests of population differentiation in Arlequin (Excoffier and Lischer 2010) to determine the degree of genetic differentiation between these seven regions. We used a Kruskal-Wallis test to determine if there was a significant difference in heterozygosity across geographic groups (McDonald 2009). We compared the overall population (all populations combined) with only individuals found in Kruger. We calculated the number of alleles and the expected heterozygosity for each of these groupings using the program Arlequin (Excoffier and Lischer 2010), and included an exact test of population differentiation to determine whether the overall population differs significantly from the Kruger population. We used a Mann-Whitney U test to determine whether heterozygosity was significantly lower in Kruger compared to the overall population.

We used the program Bottleneck v. 1.2.02 (Cornuet and Luikart 1996) to investigate recent effective population size reductions from allele data frequencies for the nine nDNA loci. For populations that underwent a recent bottleneck, this program finds a signature of excess of heterozygosity relative to expected heterozygosity (based on the observed number of alleles) since during a bottleneck, a population loses allelic diversity more rapidly than it loses heterozygosity (Cornuet and Luikart 1996). We determined the probability of a recent bottleneck using the sign test and Wilcoxon test assuming a stepwise mutation model (SMM) and a two-phase model (TPM) (proportion of SMM in TPM = 0.00, variance of the geometric distribution for TPM = 0.36) for 100,000 iterations across all 9 loci. Using the same program, we determined whether there was a deficit of rarer alleles and a modal shift in allele proportions in different allele frequency classes since these may also be indicators of a recent bottleneck (Luikart *et al.* 1998).

3.4 RESULTS

3.4.1 Mitochondrial DNA analyses

Mitochondrial DNA sequences were generated for 141 elephants. The alignment consisted of 446 nucleotide sites of which 47 were polymorphic. There were 28 distinct mtDNA haplotypes, with the individuals that carried each haplotype listed in Table B.2 (GenBank accession numbers are also listed in Table B.2). African elephant haplotypes clustered into four groups in a median-joining network (Figure 3.2), exclusive of haplotypes carried by the outgroup species (Haplotypes 24-28).

The phylogeny inferred using Bayesian analysis showed limited resolution of the relationships among haplotypes with several polytomies present (Figure 3.4). There was high support for the recovery of two distinct savannah elephant mtDNA clades, with Haplotype 15 outside of the two clades. The Bayesian analysis grouped Haplotypes 2, 3 and 4 into a separate clade, while the relationships among other haplotypes were not well resolved.

An alternative visualization of the Bayesian analysis through a cloudogram plot illustrated the uncertainty in terminal taxon relationships (Figure B.1). The cloudogram of all trees (indicated in green) and consensus trees (indicated in blue) did show Haplotype 15 as distinct from other haplotypes, but did not group the remaining savannah elephant mtDNA haplotypes into two well supported distinct clades. Other clades of high support in Bayesian analysis (e.g., Haplotypes 2, 3 and 4) are illustrated by the darker blue bands in the cloudogram.

Almost all haplotypes in the study area were identified as belonging to three of the previously identified continental mtDNA subclades described by Ishida *et al.* (2013)(Savannah-wide, Southeast-savannah and South-central subclades; Figure 3.2; Figure B.2). However, haplotypes 1 and 15, each found in Kafue and each limited to a single individual sampled in that

park (Figure 3.3), were distant from the other elephant haplotypes in the network. Haplotype 15 in particular was phylogenetically distinct in the Bayesian and cloudogram phylogenies. The most similar sequences in GenBank to H1 belonged to the North-central subclade, while H15 was closest to haplotypes within the South-Central subclade, although we did not assign these to subclades in Figure 2 given the available sequence information and lack of definitive clustering. The South-central and North-central subclades are subdivisions of the "F-clade" of African elephant mtDNA, which are believed to have originally derived from the African forest elephant (*Loxodonta cyclotis*) before being transferred to the savannah species by hybridization and backcrossing to savannah elephant males. The inter-species transfer of mtDNA is common among closely related species of elephantids and other taxa in which males and not females disperse (Petit and Excoffier 2009; Roca *et al.* 2015; Li *et al.* 2016).

3.4.2 Phylogeographic analyses of mitochondrial and nuclear DNA datasets

A Bayesian clustering algorithm implemented in Geneland identified four (K=4) genetically distinct clusters of mtDNA. Three of these mtDNA clusters were present in many national parks throughout the sampling area (see the maps of the posterior probability of belonging to a cluster in Figure 3.5), but also were not geographically contiguous. Cluster 4, which included only the single individual that carried Haplotype 1, was limited to the extreme northern part of our study area (see Table B.3 for haplotypes and number of individuals within each of the clusters). Since Geneland clusters may reflect both geographic and genetic differences, we examined the degree to which these clusters would show genetic differences by treating them as "populations". AMOVA found that most (86.06%) of the genetic variation could be attributed to variation among the "populations," with little (13.92%) variation within each of the "populations" (Table 3.1), resulting in a high Fst value of 0.86. An exact test for population

differentiation showed that cluster 4 (Haplotype 1) was not significantly different from other clusters (presumably since it represented a single sample) but all other clusters (1, 2 and 3) differed significantly, and also showed significant Fst values. The clusters identified by Geneland corresponded to clades with high support in the phylogeny (e.g. cluster 1 contains individuals with haplotype 2, 3 and 4), suggesting that the mtDNA patterns were being driven by the genetic distinctiveness among the phylogenetic clades of mitochondrial DNA. Similar to Geneland, BAPS clustering analysis of mtDNA sequence data found the best partition of data to be four populations (K=4; Log(ml)=-720.110) that were discontinuously distributed across the landscape (Figure B.3). The geographically discontinuous patterns of mtDNA (Figure 3.3) reflect the complex history of mitochondrial lineages that have persisted through time and are evident across the landscape as high mtDNA haplotype diversities in relatively small geographic space. These patterns may also reflect historical migrations that occurred when elephant distribution was continuous; such patterns for mtDNA would then have persisted due to low female dispersal among elephants. Such patterns call into question the suitability of using mtDNA to study population structure in African savannah elephants.

Geneland identified four (K=4) distinct clusters of nDNA. Each nuclear DNA microsatellite cluster was geographically contiguous and did not overlap geographically with other clusters (Figure 3.6). Cluster 1 represented a central population with individuals from Mapungubwe National Park and along the southern part of the Botswana-Zimbabwe border. Cluster 2 was found in the northern part of the study region in Chobe National Park and along the northern section of the Botswana-Zimbabwe border. Clusters 3 and 4, respectively, were centered in the northern and southern part of Kruger National Park, South Africa. However, AMOVA (Table 3.2) of the nDNA clusters found extremely low differentiation between clusters

(Fst = 0.022), with most of the variation (98%) found within clusters and little (2%) variation attributed to among cluster differences. None of the clusters was significantly different ($p>0.05$) genetically in pairwise Fst comparisons between clusters. Since Geneland clusters may reflect both geographic and genetic differences, the lack of genetic differentiation suggests that the clustering was an artifact of the geographic patterns present in the dataset, and did not reflect strong nuclear genetic differentiation among clusters. This is supported by the BAPS clustering analysis that found this nDNA dataset to conform to a single population (K=1; Log(ml)=-1889.757) (Figure B.3).

3.4.3 Comparison of Kruger elephants to other populations in southern Africa

To establish the origins of Kruger elephants, mitochondrial DNA haplotype frequencies were compared across elephants in parks north of Kruger, northern Kruger, southern Kruger, and parks south of Kruger (Figure 3.7). Although all four regions shared some haplotypes (e.g., Haplotype 14 – Figure 3.7), this comparison revealed that common haplotypes in northern Kruger National Park were also common in parks north of Kruger but not common in parks south of Kruger. By contrast, haplotypes that were common in southern Kruger National Park were also common in parks south of Kruger, but not common in parks north of Kruger (Figure 3.7). For example, haplotypes 21 and 22 were both common in north Kruger and northern parks, while being uncommon in south Kruger and southern parks. By contrast, haplotypes 12 and 2 were common in south Kruger and southern Parks, but uncommon in north Kruger and northern parks. This would seem to call into question the hypothesis that all elephants in Kruger derived from a single source population.

Finally, we would expect that Kruger National Park, if it had been subject to a founder effect followed by isolation, could show reduced nuclear genetic diversity. The elephants in our

study area were grouped into seven geographic regions and heterozygosity was determined for each of the regions. The observed nDNA heterozygosity for elephants in Kruger was higher than the value for four of the seven regions, and Kruger overall ranked fourth highest among the seven regions (Figure 3.8, Table B.4). Elephants from Hwange Region, the Botswana-Zimbabwe border Region, and Gonarezhou Region had observed heterozygosity higher than Kruger, while those in Chobe Region, Livingstone Region, and Mapungubwe Region had observed heterozygosities lower than the Kruger Region. There was low differentiation between these regions (Fst = 0.027, Table B.5) with no significant differentiation between populations (exact test of population differentiation $p>0.05$), and heterozygosity did not differ significantly between these regions (Kruskal-Wallis $K=5.62$; $p>0.5$). The Kruger population did not significantly differ genetically from the overall population (exact test of population differentiation $p>0.05$), and heterozygosity in Kruger did not differ significantly from overall heterozygosity (Mann-Whitney U test $U=33$; $p>0.5$).

We did not find evidence of a recent reduction in effective population size when considering allele data frequencies (Table B.6). Prob(H>He) is the probability that the heterozygosity (H) is larger than the average (He) under the null hypothesis, if Prob(H>He) is lower than 0.05, the null hypothesis (mutation drift equilibrium) is rejected in favor of the hypothesis of a recent genetic bottleneck. Prob(H>He) for the sign test under the SSM and TPM models was >0.5, and Prob(H>He) for both 1-tailed and 2-tailed Wilcoxon tests under the SMM and TPM models was >0.5. The mode shift test showed a normal L-shaped distribution as expected under mutation drift equilibrium.

3.5 DISCUSSION

Contrary to our hypothesis, the elephant population in Kruger did not show evidence of genetic isolation, a depleted gene pool or a strong founder effect. Genetic patterns from nDNA showed little differentiation between Kruger and nearby parks, with Kruger elephants showing levels of heterozygosity similar to those of nearby regions. Observed heterozygosity in Kruger ranked fourth highest when compared to six other populations across the study area, and heterozygosity in Kruger did not differ significantly from the overall heterozygosity of all other populations combined, and therefore did not show evidence of a depleted gene pool that typically results from genetic isolation and/or a small founder population. This is corroborated by the lack of evidence of a detectable population bottleneck, suggesting that the Kruger population did not experience a recent reduction in effective population size. Although the present study did not detect a population bottleneck, previous studies have shown that known population bottlenecks are not always detected (Whitehouse and Harley 2001; Peery *et al.* 2012). The relatively recent population extirpation followed by a rapid increase in population size (along with potential subsequent migration) may have precluded the genetic signature of a bottleneck in Kruger.

Our results suggest that the Kruger elephant population was founded by elephants from more than one geographic source, and not from a single source population as has been previously suggested (Whyte 2001). Mitochondrial DNA frequency comparisons (Figure 3.7) indicated that elephants in southern Kruger may have originated from populations south of Kruger, while elephants in northern Kruger may have migrated into Kruger from regions north of Kruger. At the least, our findings suggest that if Kruger was founded by a single source population, it has since been augmented through gene flow and migration from nearby parks either to the north or to the south of Kruger.

Phylogenetic relationships among most mtDNA haplotypes were poorly resolved by the Bayesian analyses, which is not unexpected when using a single relatively short sequence for phylogenetic inference. Future studies that aim to better determine the relationships among haplotypes should incorporate longer sequences (e.g., complete mitogenomes).

One relationship that was evident in the Bayesian analysis was the placement of Haplotype 15 outside of a clade containing the other savannah elephant haplotypes. Haplotype 15 along with Haplotypes 16-23 (South-central subclade) and Haplotype 1 (North-central subclade) (Figure B.2) are part of the F-clade, a grouping of mitochondrial DNA that is widely distributed among savannah elephants across Africa, but is believed to have originated in ancient inter-species transfer from the African forest elephant (Roca *et al.* 2015). Our findings suggest that ancient hybridization and backcrossing led to the transfer of forest mtDNA haplotypes that spread through female migration to Kafue National Park which contains Haplotype 15 as part of the South-central subclade, and also to other northern parks in the study area (e.g., Chobe National Park) that also contain haplotypes from the South-central subclade.

The discordance between mtDNA and nDNA clusters is consistent with previous reports of differences in the African elephant mitochondrial and nuclear phylogeographic patterns (Nyakaana and Arctander 1999; Archie *et al.* 2008; Okello *et al.* 2008; Ishida *et al.* 2011). Ishida *et al.* (2011) showed that genetic patterns based on nDNA and mtDNA can be quite different across the African continent. Mitochondrial DNA patterns reflect ancient female migration, with the patterns persisting due to lack of dispersal of females from core social groups (Archie *et al.* 2007; Fishlock and Lee 2013). By comparison, the male-biased dispersal typical of elephants (Nyakaana and Arctander 1999) would tend to erase geographic differences in nDNA

patterns, leading to nDNA phylogeographic patterns that may differ greatly from those present among maternally inherited mtDNA.

Management strategies for elephants in Kruger National Park have encompassed a myriad of approaches (Whyte 2001). Early management strived to increase elephant numbers in the park, later strategies aimed at decreasing elephant numbers in the park, while current strategies aim at sustainable self-regulation of population numbers through ecological processes such as resource dependent or induced dispersal and density dependent population growth stabilization (van Aarde *et al.* 1999). Our findings suggest that Kruger's elephant population has not suffered from the negative genetic consequences that may be associated with population isolation. This is likely due to historical founding events involving more than one source population, along with a rapid increase to a large population size, and possible subsequent immigration into Kruger that augmented the gene pool. Similar genetic patterns may be found in parks with similar histories to Kruger. Ultimately this study highlights the importance of migration and gene flow for maintaining genetic diversity in populations. Our findings suggest that Kruger forms part of a functional entity in which migration helped to maintain a relatively diverse gene pool. Our findings therefore emphasize the need for conservation initiatives such as transfrontier conservation areas that aim at maintaining connectivity between populations. Initiatives that maintain connectivity may provide a sustainable, self-regulating management approach for elephants in southern Africa that simultaneously upholds genetic diversity and gene flow across Kruger and nearby parks.

3.6 ACKNOWLEDGEMENTS

The projected was funded through grants to RJvA from the International Fund for Animal Welfare and the Conservation Foundation (Zambia). Elephants Without Borders (Botswana)

kindly facilitated sample collection. The Zambian Wildlife Authority, the Department of Wildlife and National Parks (Botswana), South African National Parks (SANParks) and the Department of Agriculture, Forestry and Fisheries (South Africa) sanctioned our research activities. We acknowledge the support of a team of post-graduate students that assisted with the collection of study material, and the support provided by the University of Pretoria's Sequencing Facility. Travel and research support for AdeF was provided through the Francis M. and Harlie M. Clark Research Support Grant (2016), the Harley J. Van Cleave Research Award, and the University of Illinois Graduate College Dissertation Project Travel Grant. AdeF was also supported by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number ILLU 875-952. ALR was supported by the US Fish and Wildlife Service African Elephant Conservation Fund.

3.7 DATA ACCESSIBILITY

Mitochondrial sequences have been deposited in GenBank (MF062095- MF062117).

3.8 TABLES AND FIGURES

3.8.1 Tables

Table 3.1 Analysis of molecular variance (AMOVA) of mtDNA clusters identified by the program Geneland showed high differentiation between clusters (Fst), with most of the variation attributed to differences among clusters

Source of variation	**df**	**Sum of squares**	**Variance components**	**Percentage of variation**
Among clusters	3	350.274	5.281	86.08
Within clusters	137	117.059	0.854	13.92
Total	140	467.333	6.136	100
Fixation index (Fst):	0.86076			

Table 3.2 Analysis of molecular variance (AMOVA) of nDNA clusters identified by the program Geneland showed low differentiation between nDNA clusters (Fst), with most of the variation attributed to differences within clusters

Source of variation	**df**	**Sum of squares**	**Estimated variance**	**Percentage of variation**
Among clusters	3	17.267	0.075	2
Within clusters	134	453.64	3.385	98
Total	137	470.913	3.461	100
Fixation index (Fst):	0.022			

3.8.2 Figures

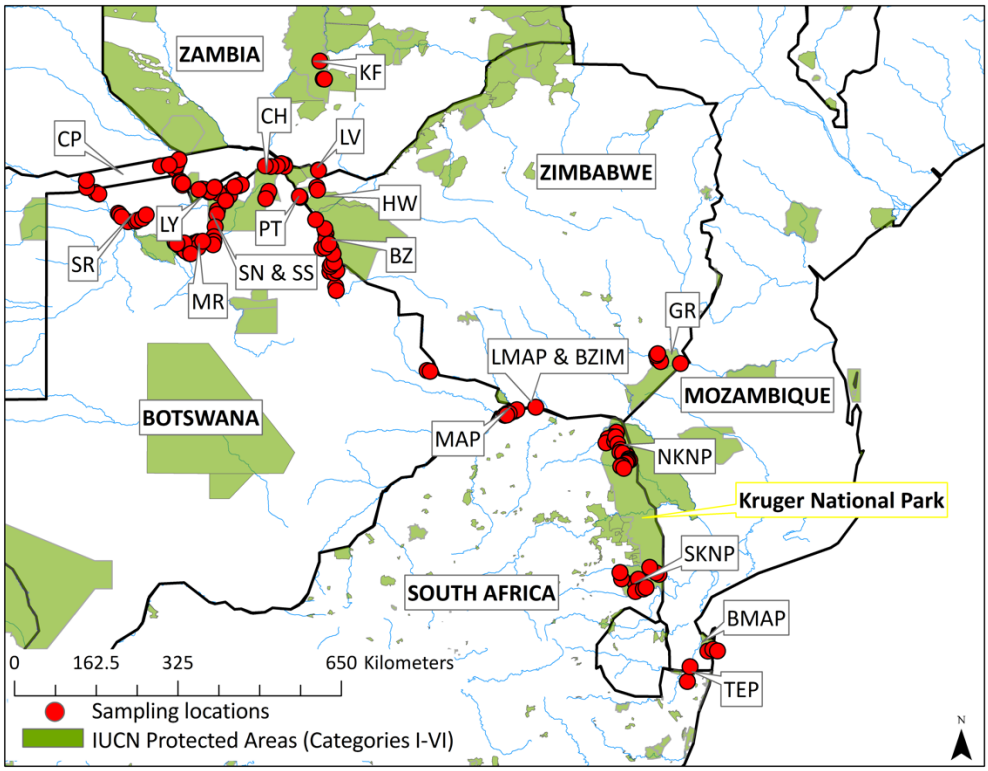

Figure 3.1 Study sites and locations where savannah elephant fecal samples were collected (circles, n=141). Study site abbreviations include SKNP and NKNP = southern and northern Kruger National Park, respectively, South Africa; MAP, LMAP and BZIM = Mapungubwe National Park (MAP) and along the Limpopo River that connects Mapungubwe and Kruger National Park (LMAP and BZIM), South Africa; TEP and BMAP = Tembe Elephant Park, South Africa and Maputo Elephant Reserve, Mozambique; KF = collected from Kafue National Park, Zambia; LV = Livingstone Town, Zambia; PT = Pandamatenga, Botswana; MR = Moremi Game Reserve, Botswana; SS and SN = south and north Savuti Game reserve, respectively, Botswana; CH = Chobe National Park, Botswana; SR = Seronga, Botswana; LY = Linyanti River, Botswana; BZ = the border of Botswana and Zimbabwe; HW = Hwange National Park, Zimbabwe; GR = Gonarezhou National Park, Zimbabwe; and CP = Caprivi region, Namibia.

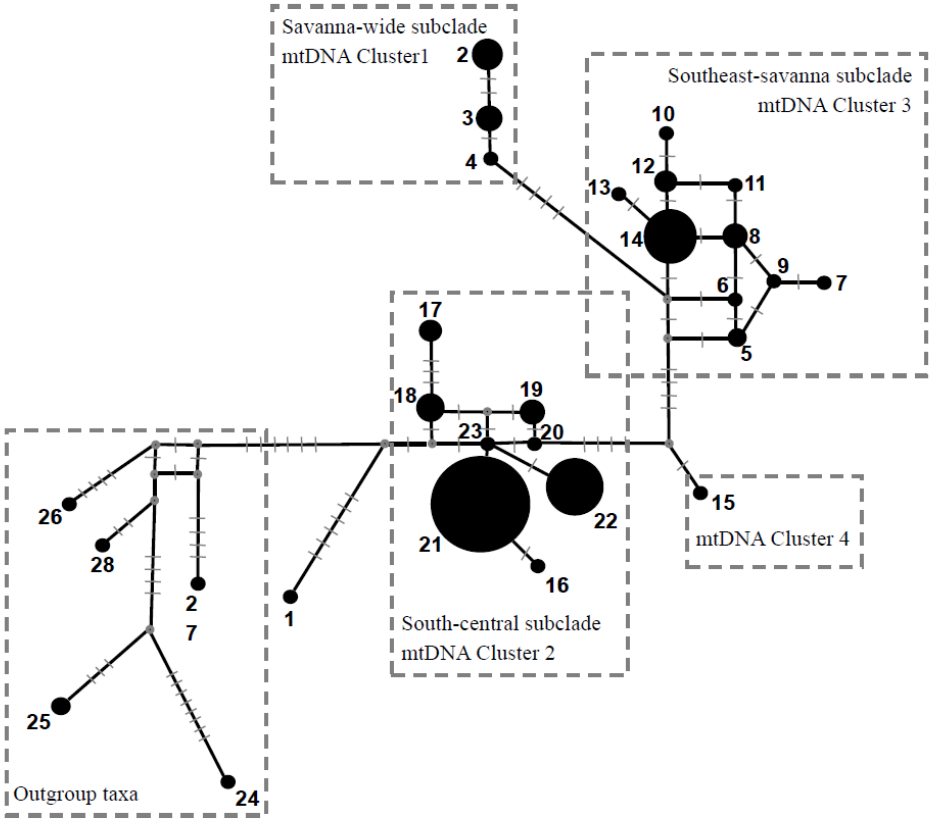

Figure 3.2 A median-joining haplotype network (based on pairwise distances) for all mtDNA sequences show separation into groups that correspond to the mtDNA clusters identified using Geneland. Outgroup sequences, groups corresponding to Geneland mtDNA Clusters 1 through 4, and previously identified mtDNA sublades (Ishida *et al.* 2013) to which haplotypes in this study were assigned are indicated in the dashed rectangles. Haplotypes H1 and H15, each found in Kafue in a single individual each, were distinctive from the other sequences and not categorized to a subclade. Branch length is proportional to the number of mutational differences (indicated as cross-hatches) between haplotypes, circle size is proportionate to the number of individuals carrying a haplotype (see Figure B.3 for an inset of the sub-clades reported by Ishida et al. 2013).

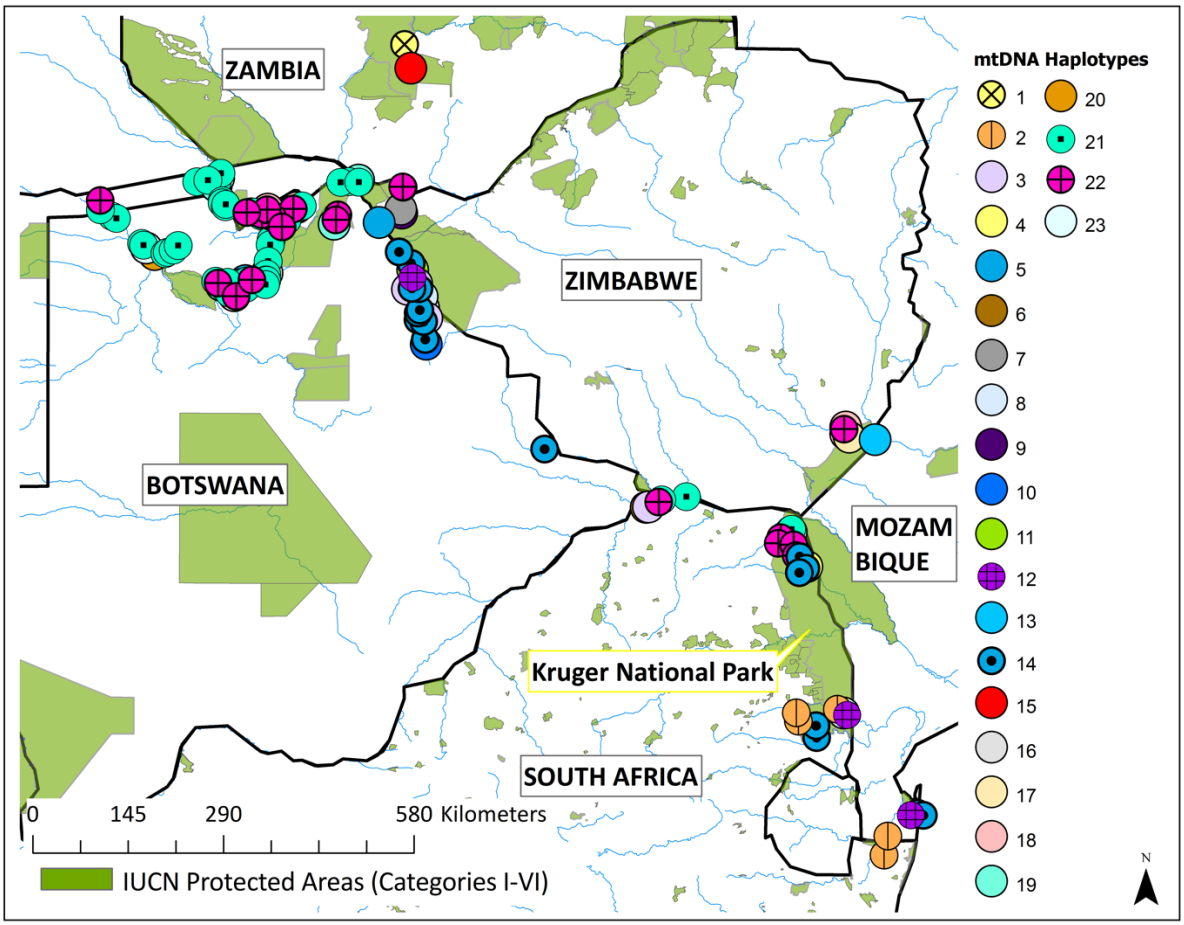

Figure 3.3 Geographic distribution of mitochondrial DNA haplotypes (color figure available online).

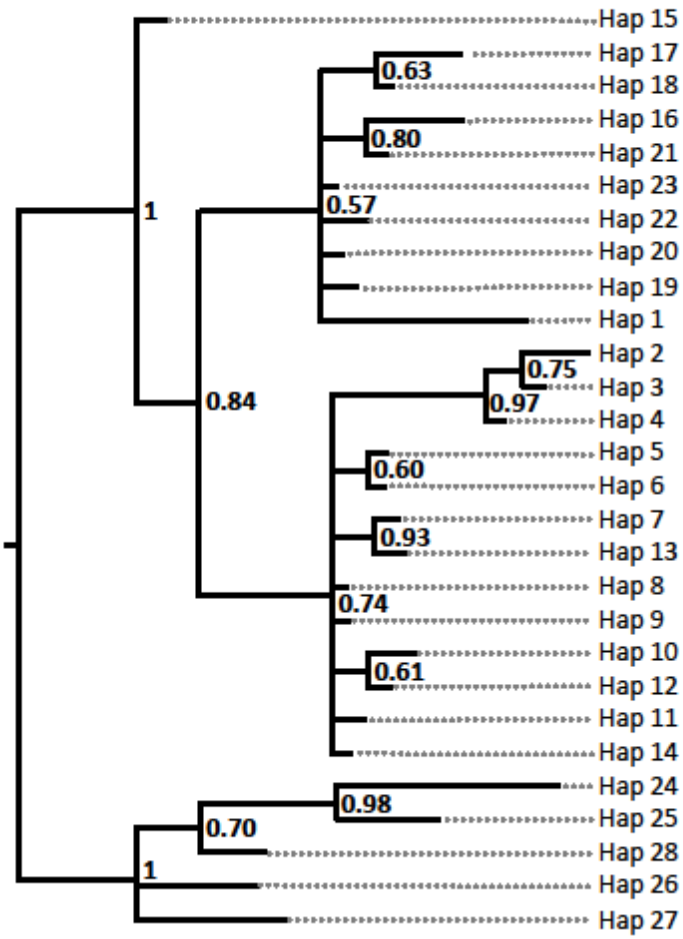

Figure 3.4 Bayesian analysis showed limited resolution of relationships between mtDNA haplotypes, where polytomies were present at several nodes. There was high support for the two distinct mtDNA clades. This topology represents a 50% majority-rule consensus tree, with node support values (posterior probabilities) shown next to their corresponding nodes.

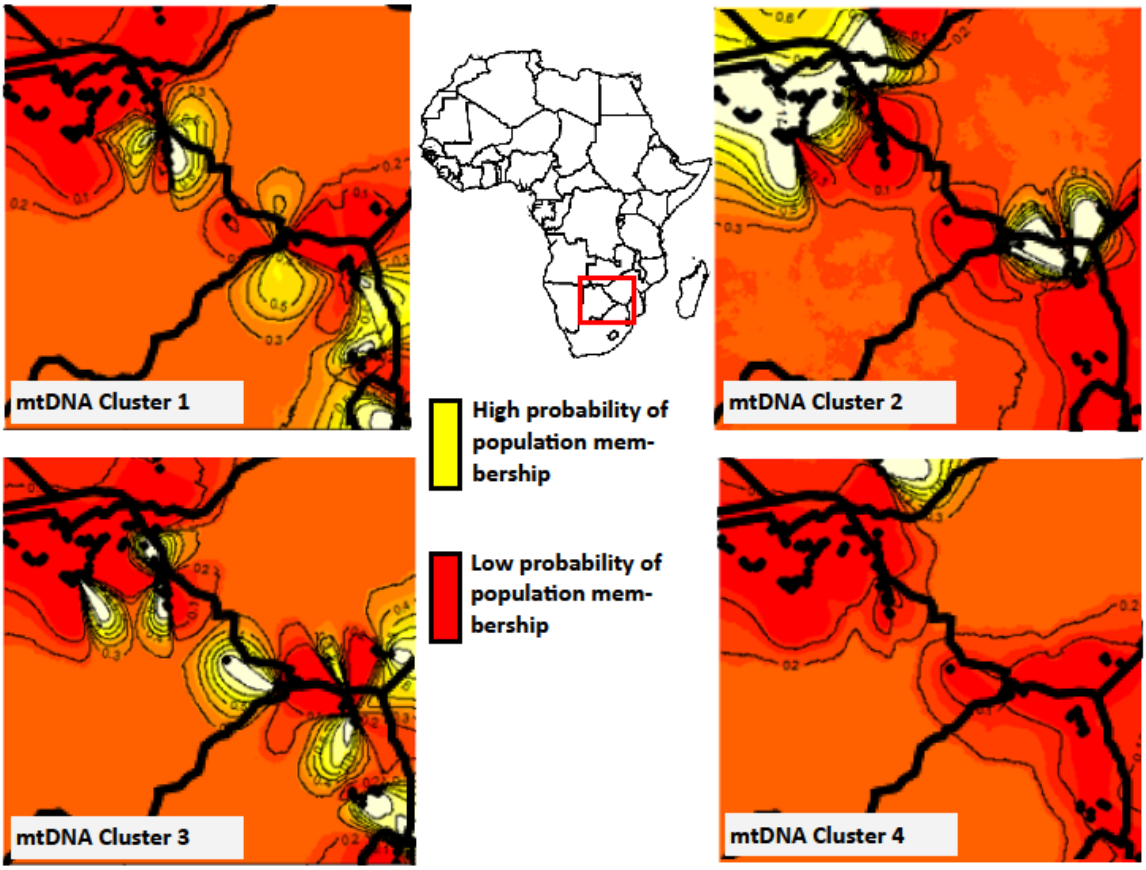

Figure 3.5 Four mtDNA clusters were identified based on a Bayesian clustering algorithm implemented in the program Geneland. Most of the "clusters" are discontinuous across geographic space, while Cluster 4 was limited to a single geographic location in the extreme north of the study region. Cluster details can be found in Table B.3. White and yellow represent areas of high probability of population membership (color figure available online).

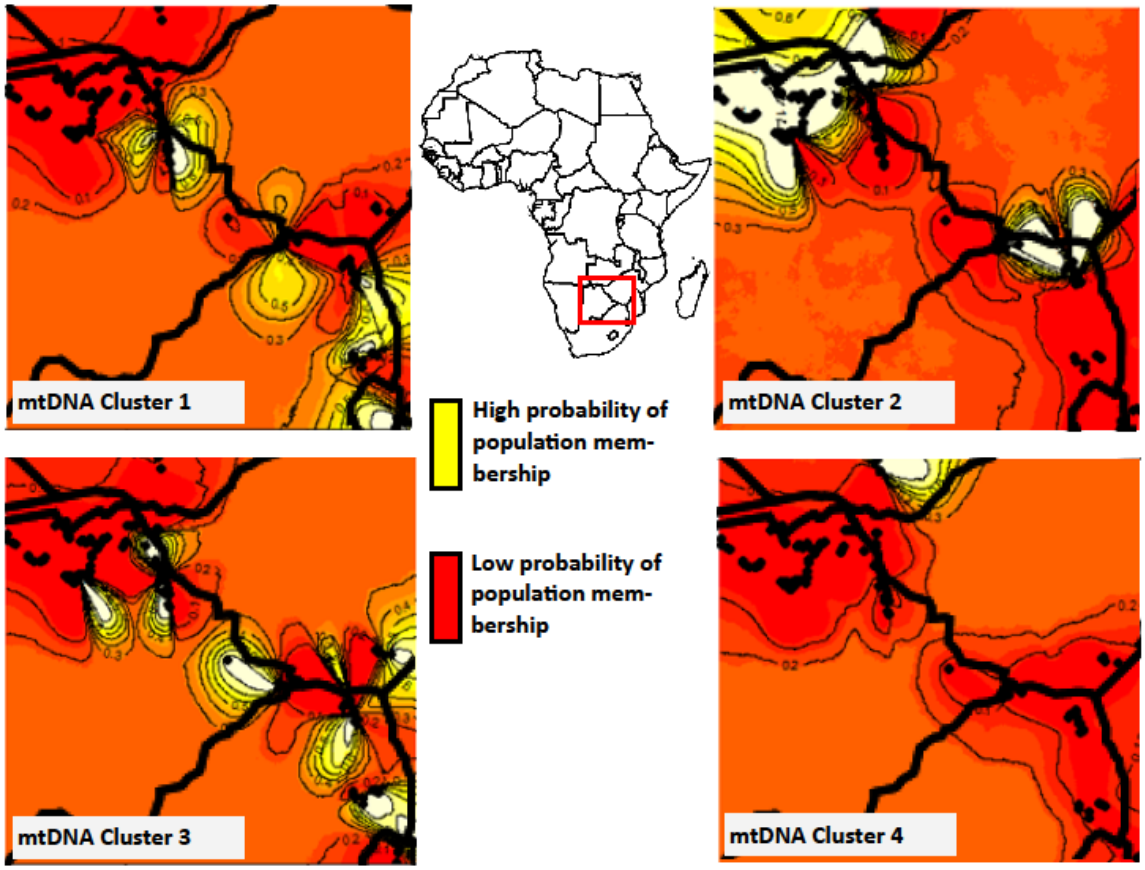

Figure 3.6 Four nDNA clusters were identified by a Bayesian clustering algorithm implemented in the program Geneland. Cluster 1 (8 individuals) falls within the center of the study area, cluster 2 (28 individuals) is limited to the extreme north, cluster 3 (24 individuals) is centered in the northern part of Kruger National Park in South Africa, and cluster 4 (9 individuals) is centered in the southern part of Kruger National Park. White and yellow represent areas of high probability of population membership (color figure available online).

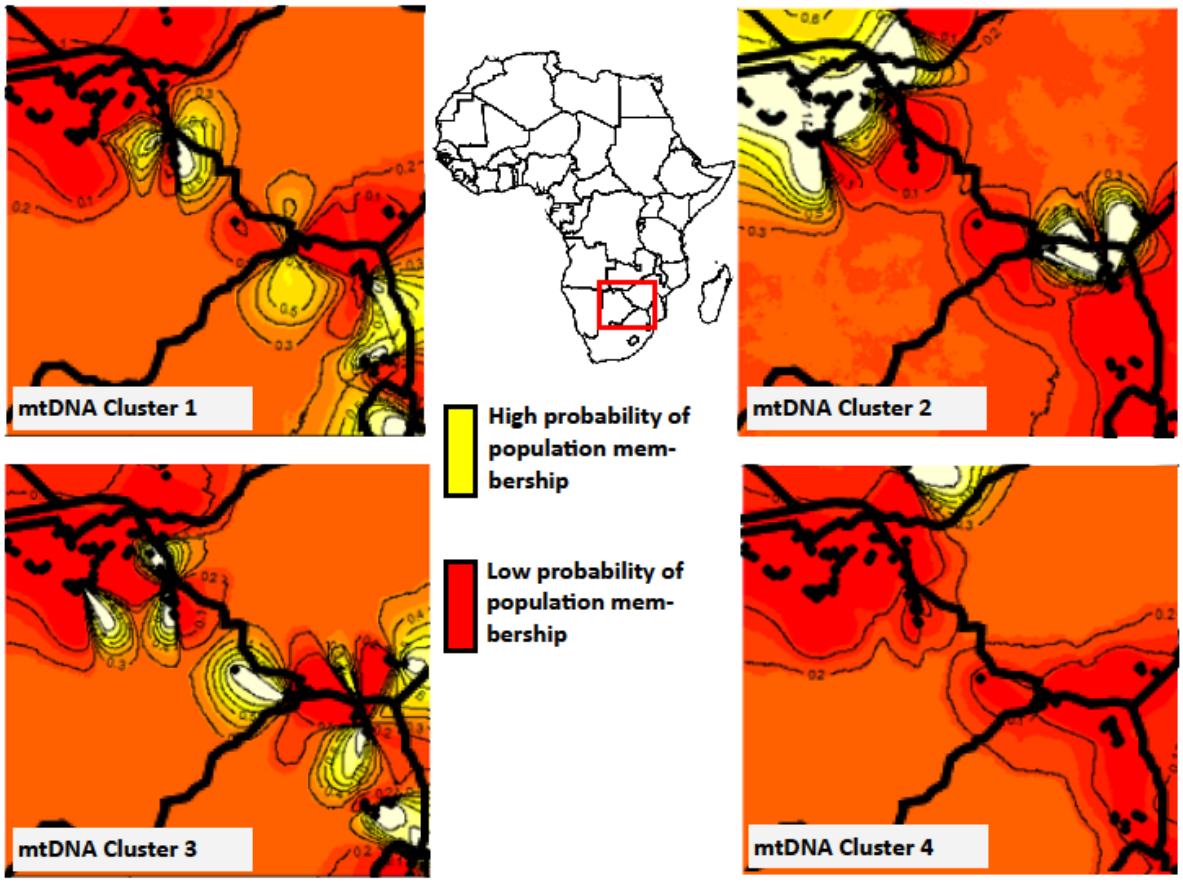

Figure 3.7 Haplotype frequencies are shown for elephants placed into four geographic groupings: north of Kruger (N=117), northern Kruger (N=12), southern Kruger (N=7), and south of Kruger (N=5). Haplotypes are color coded and shown from common (top) to rare (bottom). The haplotype color legend is shown below the bar chart, and common haplotypes in each region are indicated on the bar graph. Note the similarity in frequencies of common haplotypes, both between north Kruger and northern parks (especially haplotypes 21 and 22) and between south Kruger and southern parks (especially haplotypes 12 and 2).

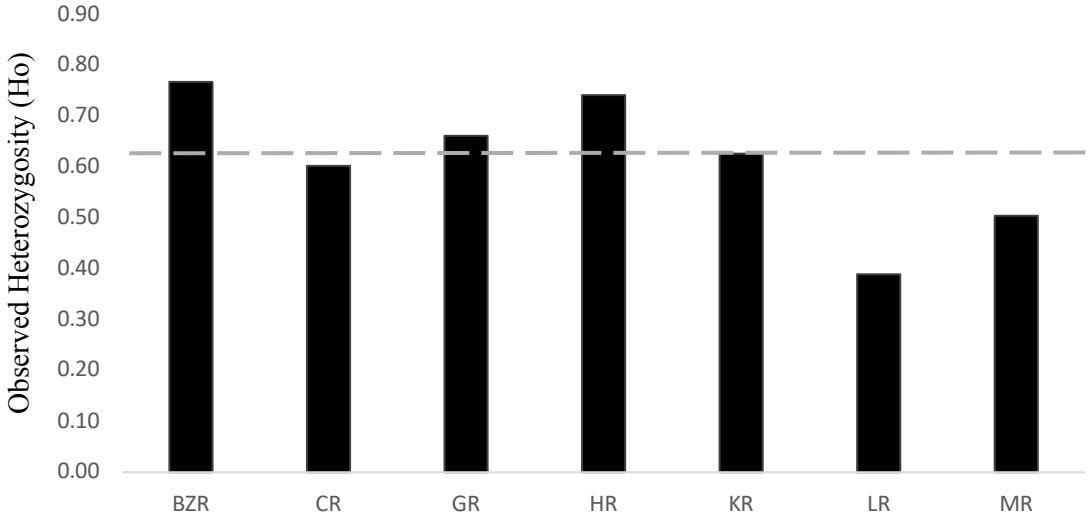

Figure 3.8 Observed heterozygosity for seven regions defined by geographic proximity and connectivity. The regions are the Botswana-Zimbabwe border Region (BZR) = BZ; Chobe Region (CR) = CH, MR, CP, SS, SN, LY; Gonarezhou Region (GR) = GR; Hwange Region (HR) = HW, PT; Kruger Region (KR) = NKNP, SKNP; Livingstone Region (LR) = LV; Mapungubwe Region (MR) = MAP, LMAP, BZIM. The dashed horizontal line indicates the observed heterozygosity of elephants in Kruger National Park, and was included to aid comparisons between regions. Locality abbreviations are defined in the main text.

CHAPTER 4: ANCIENT DNA AND ISOTOPE ANALYSES ESTABLISH PROVENANCE OF IVORY FROM A 16TH CENTURY SHIPWRECK[2]

4.1 ABSTRACT

Shipwreck cargo provides a snapshot into the past for uncovering human and non-human animal histories. On the southwest coast of Namibia, a shipwreck believed to be the *Bom Jesus*, a Portuguese trading ship lost in 1533 en route to India, yielded over 100 unworked elephant tusks of unknown provenance. Here, we use paleogenomic (ancient DNA) and stable isotope analyses to identify the source of the shipwreck ivory. Nuclear DNA identified the tusks as African forest elephant (*Loxodonta cyclotis*) rather than savanna elephant (*Loxodonta africana*) ivory. Mitochondrial DNA of the ivory cargo was characteristic of elephants from West but not Central Africa. Mitochondrial haplotype diversity indicated that the ivory derived from at least 17 different core social groups of elephants, some of which carried previously undocumented haplotypes that may have been subsequently lost due to large-scale killing of elephants for the ivory trade. Stable isotope analyses ($\delta^{13}C$ and $\delta^{15}N$) indicated that the ivory cargo derived from elephants that lived in savanna or mixed forest-savanna habitats, rather than dense forests. Such habitats surround the Guinean forest block of West Africa. Current forest elephant range in West Africa includes habitats outside the tropical forest; we find that this pattern preceded the decimation of West African savanna elephant populations in the 19th and 20th centuries. Our findings suggest that Portuguese trade networks in West Africa in the 16th century relied on communities who hunted elephants and traded ivory through networks bordering the Guinean forest block. Combining paleogenomic, archeological and historical methods provides insights

[2] Chapter 4 has been submitted for publication as de Flamingh, A., Coutu, A., Sealy, J., Chirikure, S., Bastos, A.D.S., Libanda-Mubusisi, N.M., Malhi, R.S., & Roca, A.L. Ancient DNA and isotope analyses establish provenance of ivory from a 16th century shipwreck, and is currently undergoing peer-review. The isotope experiments and analyses were conducted by A. Coutu.

into the ivory trade during the formative stages of maritime exchanges linking Europe, Africa and Asia and provides a look at the elephant historical ecology in West Africa.

4.2 INTRODUCTION

In 2008, alluvial (offshore) diamond mining on the Namibian coast north of the mouth of the Orange River uncovered the remains of a shipwreck. The ship carried over 40 tons of valuable cargo, including copper, lead and tin ingots, thousands of gold, silver and copper alloy coins, elephant tusks, and armaments (Chirikure *et al.* 2010; Werz 2010; Hauptmann *et al.* 2016; Mowa 2018). The coins were mostly of Portuguese and Spanish origin (Figure 4.1), including Portuguese 10-*cruzado* gold coins minted in 1525 at the behest of King João III (Alves 2011). These coins had a high gold content which led to them being hoarded, rather than traded; as a result, they were withdrawn by royal decree in the late 1530s (Chirikure *et al.* 2010; Alves 2011). Thus, the ship likely wrecked between 1525 and the 1530s. The ship is believed to be the *Bom Jesus*, a Portuguese trading vessel lost in 1533, reportedly "near the Cape of Good Hope" (Academia Das Ciências De Lisboa 1979; Xavier 1989; Werz 2010). This is the oldest Portuguese shipwreck discovered thus far in sub-Saharan Africa (Werz 2010). The *Bom Jesus* was a *nau* (Portuguese trading ship), designed for the India trade, one of the most lucrative and strategic commercial routes of the time. *Naus* were armed against attack and could carry large quantities of cargo and enough supplies to make the long journey through the Atlantic Ocean, around the Cape of Good Hope, and on to India (Castro *et al.* 2010).

The wreck yielded more than 100 unworked elephant tusks, which is the largest archaeological deposit of intact African elephant tusks ever found (Figure 4.1). Since only part of the ship was preserved, there may originally have been many more tusks. The tusks are of varying length and size (from 2-33 kg) indicating that elephants may have been hunted

indiscriminately, both males and females, young and old alike (Chirikure *et al.* 2010; Alves 2011). Although one may hypothesize that the tusks originated in West Africa because the Portuguese were actively trading there at that time (Chirikure *et al.* 2010; Newitt 2010; Mowa 2018), there has been no clear evidence of where they originated. The geographic provenance of these tusks, therefore, has remained unknown.

Portuguese ships explored the West African coast from 1415 (Thornton 2007). These exploratory voyages were generally made with small ships such as *caravels*, which were easy to maneuver in and out of ports along the coastline (Boxer 1991; de Castro 2005). Portuguese ships reached Cape Bojador in 1434, established a fortified trading station in Arguim (Mauritania) in 1443, reached the Senegal seaboard in 1445, then Guinea-Bissau in 1446 (Lawrence 1963; Newitt 2010). In 1462, the Portuguese founded the first permanent European settlement in West Africa on the Cape Verde Islands. Ongoing exploration reached the coast of modern Ghana in 1471, the Bight of Benin in 1475 and the mouth of the Congo in 1482 (Ijoma 1982; Garfield 1992). Also in 1482, the Portuguese built a fort at an established trading post in West Africa, São Jorge de Mina, or Elmina, Ghana, on the south-western edge of the Dahomey Gap (Blake 1942; Fritze 2002) (Figure 4.2). The Dahomey Gap comprises a mosaic of drier-land vegetation types, providing a corridor through dense forest between the coast and the interior (White 1983). This vegetation distribution has existed for thousands of years (Sowunmi 2004). Coastal settlements near the Dahomey Gap expanded as Elmina became an important entrepôt for Atlantic Ocean ships from the 15th century.

By the early 16th century, the Portuguese had four major fortified trading bases along the West African coast: Arguim (Mauritania), Santiago Island (Cape Verde), Elmina (Ghana) and São Tomé Island, shown on Figure 4.2 (Lawrence 1963; Boxer 1991; Newitt 2010). Here, the

Portuguese crown attempted to control Atlantic trade (primarily in gold and slaves) by linking into already established African trade networks, excluding other European nations and prohibiting unauthorized trading by private Portuguese merchants (Mark and da Silva Horta 2013).

Raw and carved ivories were exported from the West African coast to Portugal from the mid-15th century in exchange for European goods such as textiles and metals (Thornton 1998; Afonso and da Silva Horta 2013). Ivory objects such as salt cellars, spoons, and olifants (hunting horns made of elephant tusk) were carved by West African artists as part of an indigenous artistic tradition, adapted in these 'Luso-African ivories' to include European motifs in order to appeal to the European market (Afonso and da Silva Horta 2013; Mark and da Silva Horta 2013). Elephant ivory also formed part of gift exchanges between West African ambassadors and the Portuguese king (Afonso and da Silva Horta 2013). Ivory exported to Portugal during the 15th – 17th centuries originated from three main regions along the West African coast: Senegambia, the Gulf of Guinea, and the Loango coast south to Angola (Soares 2017) (see Figure 4.2). Islands such as Cape Verde and São Tomé became important centers of trade: ivory (and other goods) were trans-shipped to the islands in order to be counted, weighed, and then sent to Casa da Índia in Lisbon, which was the central clearing house for African and Indian imports to Portugal (Garfield 1992; de Castro 2005; de Alencastro 2007; Soares 2017). Because the Great Lisbon earthquake and fire of 1755 destroyed the archives of the Casa da Índia, identification of the shipwreck as the *Bom Jesus* relied on other records (e.g. Academia Das Ciências De Lisboa 1979; Xavier 1989). Centralized loading of outgoing long-distance trading vessels in Lisbon enabled tight control of the valuable cargo. *Naus* designed for the India trade sailed south-westwards from Portugal, stopping in Cape Verde for water and supplies, then across the

Atlantic and south-eastwards on the trade winds (de Castro 2005; Newitt 2010). Between 1531 and 1540, 80 vessels including the *Bom Jesus* sailed from Lisbon to Asia on the India route (Schwartz SB 2007). Due to the strong current east of the Gulf of Guinea, ships on the India route typically did not tack along the West African coastline on their departure from Cape Verde (de Castro 2005; Newitt 2010). This would preclude the *Bom Jesus* having picked up ivory along the West African coast on the outward-bound journey to the East and instead suggests the ivory on the ship probably came from a stockpile from earlier trade.

Large-scale extraction of raw ivory became a major part of Atlantic commerce, with Portuguese ships supplying an African commodity to India via the maritime routes of the Indian Ocean. This external, global demand for ivory relied on pre-existing West African trade networks, yet very little is known about the geographic distribution of the elephants which were hunted for the trade, the communities who moved this ivory across these regions, or how local networks were transformed by incorporation into Atlantic commerce. This study, therefore, uses a combination of ancient DNA and stable isotope analyses to determine the provenance of the tusks recovered from the *Bom Jesus* shipwreck to shed light on the trade in ivory, which linked hunters, craftsmen, traders of West Africa and Portuguese merchants in the 16th century.

Molecular methods can be used to trace the geographic provenance of elephants by using extracted DNA. Mitochondrial DNA (mtDNA) can establish the provenance of elephants because it is transmitted only by females. Female elephants are matrilocal and usually remain with their natal social groups throughout their lives in a relatively restricted geographic range (Archie *et al.* 2007; Fishlock and Lee 2013). African elephant mtDNA haplotypes group into 8 subclades, each of which is geographically restricted (Ishida *et al.* 2013) (Figure 4.3A). Each subclade includes a large number of distinctive DNA sequence haplotypes, most of which have

been reported only from elephants within a single country or a single locality within a country (Ishida *et al.* 2013).

African elephants comprise two genetically distinct species, the African savanna elephant (*Loxodonta africana*) and the African forest elephant (*Loxodonta cyclotis*) (Grubb *et al.* 2000; Roca *et al.* 2001). Today, *L. cyclotis* occur mostly in densely forested environments (Maisels *et al.* 2013). Exterior features of the tusks such as morphology and size vary within and between species and between the sexes; for example, large tusks have been found in both savanna and forest elephant bulls (Sukumar 2003). A molecular approach is thus necessary to determine the species of elephant from which the shipwreck tusks derived. Historical hybridization and backcrossing between elephant species has led to introgression of forest elephant mtDNA into savanna elephants (Roca *et al.* 2015). Therefore, nuclear DNA (nDNA), which shows little or no introgression between the species, was examined to ensure accurate species identification (Ishida *et al.* 2011).

Carbon and nitrogen stable isotope ratios were measured in the shipwreck ivory to determine the diets and habitats of the elephants from which the tusks derived. African elephants are mixed feeders, consuming a variety of plants, including trees, grass, bark, and fruit. African elephants living in dense forest environments primarily consume C_3 plants, while elephants living in more open shrub or grassland environments consume a mixture of C_3 shrubs and trees and C_4 grasses. The mean $\delta^{13}C$ for modern C_3 plants is -28.77 ± 2.68‰ (n=3478), and for modern C_4 plants is -12.90 ±1.52‰ (n=137) (Cornwell *et al.* 2016). The stable carbon isotope ratios measured in ivory originate from the plant food the elephant consumed, which in turn is an indication of the habitats that had been inhabited by the elephants. The $\delta^{13}C$ and $\delta^{15}N$ values for collagen extracted from ivory (dentine) are similar to those for bone collagen: Published

$\delta^{13}C_{collagen}$ for wild African elephants range from -27‰ for pure C_3 feeders in deep forests to -11‰ for savannah elephants consuming substantial quantities of C_4 grass (Van der Merwe *et al.* 1990; Cerling *et al.* 2007; Codron *et al.* 2012; Coutu *et al.* 2016; Ziegler *et al.* 2016).

Variation of ^{15}N in elephants is driven primarily by nitrogen cycling in the soil, which is strongly influenced by moisture availability (Amundson *et al.* 2003; Aranibar *et al.* 2004; Murphy and Bowman 2006, 2009). Published $\delta^{15}N_{collagen}$ values for wild African elephants range from 2 to 17‰ in moist to arid areas, respectively (Heaton *et al.* 1986; Van der Merwe *et al.* 1990; Vogel *et al.* 1990; Ishibashi *et al.* 1999; Cerling *et al.* 2007; Codron *et al.* 2012; Coutu *et al.* 2016; Ziegler *et al.* 2016). If the elephants from the shipwreck ivory cargo originated from a range of diverse habitats with different rainfall regimes, we would expect to see a similarly large range in the $\delta^{15}N$ values of the shipwreck ivory.

We thus used paleogenomic (ancient DNA) methods to identify species of elephant and the broad geographic range (West, Central, Southern and/or East Africa) from which tusks were harvested, and stable isotope analysis to identify the vegetation and habitat in which the elephants lived and fed. Integrating paleogenomic analyses and stable isotope measurements with historical records allowed us to infer the origins of shipwreck ivory with a resolution that would not have been possible using any one of these approaches on their own.

4.3 RESULTS

4.3.1 Paleogenomic results

DNA was successfully extracted from 44 of 62 (71%) of the shipwreck ivory samples for which sufficient material was available. A ~436 bp region of the D-loop mtDNA control region was amplified for all 44 shipwreck ivory samples. In order to compare the shipwreck ivory sequences to all 8 of the African elephant mtDNA subclades previously reported (Ishida *et al.*

2013), a median-joining (MJ) network was constructed (Figure 4.3B) that included sequences from the shipwreck ivory and from 24 previously published reference samples (Table C.1) representative of the 8 subclades. The shipwreck ivory samples grouped only with the West-central and the Western subclades of mtDNA, and did not group with any of the other subclades (Figure 4.3). Out of the 44 ivory DNA samples, 23 grouped with the Western mtDNA subclade, which is carried only by elephants from West Africa. Thus these 23 tusks were determined to be from West Africa. The other 21 sequences grouped within the West-central subclade, which is carried by elephants in both West and Central Africa (Fifure 4.3A). The shipwreck ivory had a total of 17 distinct mtDNA haplotypes (Figure 4.3B). Female elephant philopatry and the lack of dispersal from core natal social groups suggest that the ivory was derived from at least 17 core social groups ("herds") (Archie *et al.* 2007; Fishlock and Lee 2013).

Although subclades of mtDNA have broad geographic distributions across Africa, the distribution of each distinct haplotype sequence within a subclade may be much more limited geographically. For this reason, the shipwreck ivory mtDNA haplotypes were compared to 37 previously published mtDNA sequences which are also known to be within the West-central (n=29) and Western (n=8) mtDNA subclades (Table C.2). A 336 bp alignment of overlapping sequences was used to generate a median joining network. Many of the shipwreck ivory sequences grouped with haplotypes carried only by elephants from West Africa. While some shipwreck ivory sequences were also carried by elephants from both West and Central Africa, in no case did the shipwreck sequences match haplotypes carried exclusively by elephants in Central Africa (Figure 4.3C). Three of the ivory haplotypes were shared by only West African elephants (Figure 4.3C – black arrows), and only a single ivory haplotype was shared with both West and Central African elephants.

To further examine the origin of the shipwreck ivory, we generated the same ~436 bp region of the D-loop from 16 geographically-referenced elephant samples collected across Africa in the late 20th century (Table C.3) from a reference collection at the University of Cape Town, South Africa (UCT). These newly sequenced 20th century geographically-referenced samples showed that in West Africa, the West-central and Western mtDNA clades extended further west geographically than has been previously reported, with both clades detected in Liberia and Sierra Leone. In a median-joining network, the shipwreck ivory haplotypes grouped with the haplotypes of the newly sequenced 20th century West African elephants (Figure C.1). A West African origin for all of the shipwreck ivory is further supported by the complete absence among the ivory samples of haplotypes from the North-central, East-central or South-central subclades of African elephant mtDNA, which are common in elephants from Central Africa (Figure 4.3A) but absent from West Africa. Overall, the evidence strongly supports a West African origin for the ivory.

For nine of the shipwreck ivory samples, we assembled complete mitogenomes, with a range of 3-36X average coverage for individual mitogenomes and ~13X average coverage across all mitogenomes (Table C.4). These were compared to 11 previously published mitogenomes (Table C.5). The ancient shipwreck ivory mitogenomes grouped with mitogenomes from individuals from West Africa (Sierra Leone and the Ivory Coast) in a maximum likelihood tree (Figure 4.3D). The shipwreck ivory DNA showed damage patterns typical of ancient DNA (Figure C.2) with increased nucleotide misincorporations towards the terminal ends of the DNA molecules (Briggs *et al.* 2007). Although present, the ivory DNA damage patterns were minor, likely due to the good preservation of the ancient ivory samples. The genomic libraries used to assemble complete mitogenomes were constructed using multiple DNA extractions from the

same samples as those used for the shorter D-loop analysis. For all DNA extractions per sample, the complete mitogenome and D-loop analyses resulted in consistent mtDNA subclade assignments for the sample.

Among African elephants, haplotypes from three mtDNA subclades (Northern-savanna, Savanna-wide and Southeast-savanna subclades) have been detected only among savanna elephants (see Figure 4.3A). The complete absence of haplotypes from these three subclades among the shipwreck ivory suggested that the tusks were from forest and not savanna elephants. However, mtDNA from the Western and West-central subclades is known to have introgressed from forest into savanna elephants, so that nuclear DNA markers are more appropriate for genetically identifying the two species of African elephant. We examined SNPs within 3 short unlinked chromosome segments that are known to be fixed between the two species (Ishida *et al.* 2011). In each case, the nucleotide at the SNP matched the character state diagnostic of the African forest elephant and in no case did it ever match that of the African savanna elephant (Table C.6).

4.3.2 Stable carbon and nitrogen isotope results

Dentine collagen was successfully isolated from 97 of 100 shipwreck ivory samples (Table C.7). $\delta^{13}C$ values ranged from -22.2 to -17.1‰, with a mean of -20.4±1.2‰ (Figure 4.4). The most negative values were from elephants that consumed purely C_3 diets; more positive values reflected some C_4 grass consumption. $\delta^{15}N$ values of 97 shipwreck ivory samples ranged from 4.8 to 9.0‰, with a mean of 6.8±0.8‰ (Figure 4.4). Such values typically derive from mesic terrestrial environments. Taken together, the range of values seen here indicate that these tusks derive from elephants living in a range of habitats with different rainfall regimes and vegetation.

For four tusks, we had samples large enough to analyse a series of seasonal growth increments, in order to assess within-individual dietary variation over time (Figure C.3). We conducted this analysis to compare the within-tusk variation to the variation between individuals found in the single averaged measurement from each tusk, shown in Figure 4.4. $\delta^{13}C$ values measured in each of the four tusks ranged from -24.0 to -17.1‰ and $\delta^{15}N$ values from 5.6 to 8.3‰. Only one tusk had within-tusk measurements which indicated a cyclical pattern of increasing and decreasing $\delta^{13}C$ values over a series of 23 growth increments. This pattern could indicate switching between C_3 browse and C_4 graze in dry and wet seasons, respectively, as documented in stable isotope values from serially-sampled tissues (ivory and tail hair) in modern savanna elephants from eastern and southern Africa (Cerling *et al.* 2009; Codron *et al.* 2012, 2013; Coutu 2019). Overall, the range of values found in the four tusks that were serially sampled are typical of elephants that live in savanna environments and feed on both C_3 and C_4 plants. These values thus do not reveal substantial changes in consumption patterns or habitat over multiple years when the tusk was growing.

4.3.3 Sample preservation

The shipwreck and its contents are very well preserved (Figure 4.1; Chirikure *et al.* 2010), and we were able to extract DNA successfully from more than 70% of the ivory samples tested, and collagen from 97%. The high quality of molecular preservation in the shipwreck ivory may be due to the ivory having been consistently waterlogged since the wreck in the cold water of the Benguela Current, which carries cold Antarctic water northwards up the west coast of Africa. Cold temperatures are suggested to be effective at preserving ancient DNA in historical samples (Höss *et al.* 1996).

4.4 DISCUSSION

The ivory sampled from the *Bom Jesus* shipwreck originated from West African forest elephants (*Loxodonta cyclotis*). The tusks were identified as being from this species using nuclear loci with fixed differences between the two African elephant species, which was also consistent with the absence of tusks from three mitochondrial subclades found only among savanna elephants. There were also overlapping lines of evidence establishing that the provenance of the ivory was West Africa. Twenty-three of the shipwreck tusks had mtDNA that matched the Western mtDNA subclade, which is present only in West Africa (Figure 4.3A, B). While haplotypes from the remaining 21 shipwreck tusks grouped with the West-central subclade found in both West and Central Africa, these sequences matched geographically-referenced samples from West Africa and not Central Africa (Figure 4.3C). The complete absence of haplotypes from other mtDNA subclades further supported a West African origin for the tusks. Sequences from the North-central, East-central and South-central subclades would almost certainly have been present in the tusks had the ivory been harvested in Central Africa (Figure 4.3A), but these were not detected.

The $\delta^{13}C$ and $\delta^{15}N$ values of the shipwreck ivory cluster most closely with modern elephants from environments with mosaic vegetation encompassing both shrub and wooded savanna in Benin, Burkina Faso, Chad, Central African Republic, Niger and some elephants from Angola and South Africa (Figure 4.4; Table C.8; CILSS - Comité Permanent Inter-états de Lutte contre la Sécheresse dans le Sahel 2016). Shipwreck ivory $\delta^{13}C$ values are more positive than those of modern elephants from Cameroon, Democratic Republic of the Congo, and Liberia, which have ^{13}C-depleted values reflecting pure C_3 diets in deep forest/rainforest habitats with continuous tree canopy (Van der Merwe and Medina 1991; Cerling *et al.* 2004; Bonafini *et al.* 2013; CILSS - Comité Permanent Inter-états de Lutte contre la Sécheresse dans le Sahel 2016).

Shipwreck samples with the most positive $\delta^{13}C$ values fall within the range of modern comparative samples from open or shrub savanna environments, where elephants consume substantial proportions of C_4 grasses (Cerling *et al.* 2009; Codron *et al.* 2012; Coutu *et al.* 2016). $\delta^{15}N$ values fall in the lower to middle part of the range documented for African elephants, encompassing the values for elephants living in habitats in Angola, Benin, Burkina Faso, Chad, Central African Republic and Niger. They also overlap with values for ivory from East African coastal forests (Coutu *et al.* 2016), which have discontinuous canopies and patches of open grassy vegetation (White 1983). The shipwreck ivory $\delta^{15}N$ values do not cluster with those of elephants from Angola, Namibia or South Africa, where high $\delta^{15}N$ values (above 10‰) result from open, arid grassland savanna environments, nor with elephants from the Democratic Republic of the Congo, which have high $\delta^{15}N$ values due to living in dense forests (Cerling *et al.* 2009; Coutu *et al.* 2016).

Taken together, the $\delta^{13}C$ and $\delta^{15}N$ values of the shipwreck ivory extend over 46% of the total variation in $\delta^{13}C$ and 49% of the total variation in $\delta^{15}N$ measured in the geographically-referenced African elephants shown in Figure 4.4. The cargo of tusks was therefore likely sourced from elephants that lived in many different habitats, but not from deep forests or arid environments. This is consistent with the mtDNA evidence that the elephants derived from at least 17 core social groups ("herds"). The assignment of the ivory to this large number of elephant lineages from different habitats suggests that the ivory loaded onto the *Bom Jesus* was sourced from stockpiles in Lisbon, where the ship was loaded for the voyage to India (Soares 2017).

In West Africa, the historic range of savanna elephants was likely continuous across the Sahelian/Sudanian savanna habitat belts north of the Guinean forest block (Frade 1955). The

current distributional range of forest elephants includes habitats both inside and outside the tropical forest (Tchamba and Seme 1993; Groves and Grubb 2000; Grubb *et al.* 2000; Mondol *et al.* 2015). Mondol et al. (2015) found that elephants with forest elephant genotypes occur in savanna landscapes in the Pendjari-Arli complex in West Africa, while Groves and Grubb (2000) report skull morphologies typical of forest elephants from some West African savannas. In West Africa, the dense tropical lowland rainforests of the Guinean forest block are currently surrounded by dry forest and thicket, and a mosaic of forest and savannah (Figure 4.2) (White 1983). For example, the Dahomey Gap on the eastern edge of the Guinean forest block, which existed long before the 16th century (Sowunmi 2004), is a mosaic of savannah and drier type lowland rain-forest. Such secondary forests, coastal savannahs and derived savannahs were present in coastal regions on either side of the Guinean forest block during the 16th century and may have been utilized by forest elephants during that time. Before our analyses, the recent distribution of forest elephants in West Africa outside of tropical forest habitats could be attributed to the decimation of savanna elephants in West Africa in the 19th and 20th centuries (Michelmore *et al.* 1994; Maisels *et al.* 2013). Our combined genetic and isotope results suggest that utilization of savanna habitats by forest elephants in West Africa preceded the decimation of savanna elephants, and dates back to at least the 15th century.

Contemporary West African forest elephant populations may have limited genetic diversity compared to historical populations. Only two of ten Western subclade mtDNA haplotypes present among the 16th century tusks have been reported among contemporary populations (Figure 4.3C), likely reflecting the reduction of West African elephant range by 93% in the last century (Roth and Douglas-Hamilton 1991). This is of conservation concern because decreases in genetic diversity have been associated with negative demographic and physiological

outcomes, such as expression of deleterious alleles, reduced reproductive fitness and increased risk of population extirpation (Allendorf *et al.* 2013).

4.5 CONCLUSION

This interdisciplinary study uses paleogenomic and stable isotope approaches to analyze a large cargo of elephant tusks recovered from a 16th century shipwreck, providing information on the species of elephant, their geographic provenance, and habitat characteristics of the landscape in which they lived. The paleogenomic results determined that the ivory derived from forest elephants from West Africa, while the stable isotope data showed that the elephants lived outside of the deep tropical forest. This study provides new insights into 16th century Portuguese maritime routes, involving major trading posts such as São Jorge de Mina, or Elmina (Fritze 2002). Based on combined genetic and isotope analyses, the elephants hunted for the ivory cargo of the *Bom Jesus* originated from forest elephants from habitats outside of the Guinean forest block. The tusks are evidence for exploitation of raw ivory for export from West Africa via Portuguese vessels, to supply the Indian Ocean rim with a sought-after commodity.

Our study demonstrates the value of an interdisciplinary approach that incorporates archeological, paleogenomic and conservation genetic methodologies, combining genetic and isotopic analyses of ancient ivory with historical information. In addition to identifying the origins of tusks from a 16th century shipwreck, we have been able to reconstruct long-term histories of elephant populations relevant to conservation genetics. With a resolution not possible using any single approach, our interdisciplinary methodologies enable examination of patterns of ivory acquisition and circulation during the formative stages of maritime trade that linked Europe, Africa and Asia. Future studies on historical ivory and shipwreck cargoes should

incorporate interdisciplinary approaches such as ours for understanding the historical ecology of an endangered species, as well as to reconstruct the history of maritime and terrestrial commerce.

4.6 METHODS

4.6.1 DNA extraction

DNA was extracted from 62 ancient ivory samples using an ancient DNA extraction protocol previously developed and optimized by the Malhi ancient DNA laboratory (Cui *et al.* 2013) at the Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign (UIUC). All shipwreck ivory DNA extractions were conducted in this facility in a laboratory dedicated to the analysis of ancient DNA. DNA was extracted from 20th century historical reference samples from the University of Cape Town (UCT) collection in a dedicated Bio-Safety Level 2 laboratory at the University of Pretoria (UP), South Africa. The UP laboratory was decontaminated with DNA-off prior to the initiation of the project, and the entire laboratory was UV sterilized on a daily basis using a UV ceiling light for at least 1 hour. Using swabs, we sampled multiple surfaces in the laboratory and none of the surface samples yielded PCR amplicons for elephant mtDNA. No samples other than the 20th century historical reference samples were processed in this laboratory for the duration of this project. PCR amplification of both the shipwreck ivory and the 20th century historical reference samples was carried out in laboratories that were isolated from other laboratories in which DNA extractions were carried out. To ensure that external or cross-sample contamination was avoided, extraction and PCR-negative controls were included with each round of sample processing, and not more than 8 samples were processed at any one time.

4.6.2 mtDNA D-loop sequencing and analyses

A 436 bp fragment was amplified for 44 of the 62 ivory samples, and for 16 twentieth-century geographically-referenced samples from the UCT collection, using published mtDNA D-loop primers (Nyakaana and Arctander 1999). These sequences were compared to those of eight previously reported African elephant mtDNA subclades (Ishida *et al.* 2013). We downloaded three 4258-base pair (bp) reference sequences for each of the eight mtDNA subclades from the original dataset used by (Ishida *et al.* 2013), aligned our shorter 436 bp sequences to the longer reference sequences using the program MUSCLE (Edgar 2004), and trimmed the sequences so that only regions present in both the reference sequences and the ivory sequences were included in the final alignment. The alignment was then used to construct a median-joining network in the software POPART (Population Analysis with Reticulate Trees) (Leigh and Bryant 2015) that compared shipwreck ivory to mtDNA subclade reference sequences (Figure 4.3B), to novel twentieth-century geographically-referenced and published mtDNA sequences from West and Central African elephants (Figure C.1).

A geographically-referenced database was compiled for haplotypes within the two mtDNA subclades that the ivory grouped with (Western and West-central mtDNA subclades) by downloading previously reported D-loop mtDNA elephant sequences from GenBank and coding them to reflect the geographic origin (West Africa or Central Africa) of those elephants. We aligned (Edgar 2004) the 436 bp ancient ivory sequences to the geographically-referenced database (336 bp), and trimmed the data so that only overlapping regions were present in the final alignment. The alignment was then used to infer a pairwise-distance based median-joining network (Leigh and Bryant 2015) (Figure 4.3C). It is important to note that some studies reporting reference sequences failed to list the frequency at which each haplotype was observed. However, the current study sought to determine only whether the ivory matched geographically-

referenced haplotypes from West or from Central Africa, and thus the frequency at which haplotypes were observed in populations was not of concern.

4.6.3 Mitogenome sequencing and analyses

Genomic libraries were constructed for nine ancient ivory samples using the NEBNext® Ultra II™ DNA Library Prep kit and NEBNext® Multiplex Oligos (Unique Dual Indexes) for Illumina®. Libraries were pooled and shotgun sequenced on a HiSeq 4000 platform at the UIUC Core Sequencing Facility. Reads were de-multiplexed and trimmed using AdapterRemoval (Lindgreen 2012) to have a minimum sequence length of 25 bp. Reads were aligned to the assembled African elephant genome (*Loxodonta africana* assembly Loxafr 3.0) and to a published forest elephant mitogenome (Brandt *et al.* 2012a) using bowtie2 (Langmead and Salzberg 2012) with the local alignment option, and capping fragment length at 1000 bp. Aligned sequences were transformed to BAM format in SAMtools v. 1.1 (Li *et al.* 2009). Using SAMtools, BAM files were filtered to remove unmapped reads and reads with a quality score less than 30, were sorted and indexed, and PCR duplicates were removed. Consensus sequences were generated from the de-duplicated alignment files in Geneious R7 (www.geneious.com). Ancient DNA damage patterns were verified by aligning trimmed reads to the African forest elephant mitogenome with BWA (Li and Durbin 2010) and quantifying damage in mapDamage2 (Jónsson *et al.* 2013) using a fragment size of 70 bp. All bioinformatic analyses were performed using the Biocluster2 supercomputer of the Carl R. Woese Institute for Genomic Biology.

Consensus sequences of ancient ivory mitogenomes were compared to 11 previously published African elephant mitogenomes (Table C.6) by inferring a maximum likelihood (ML) tree in RaXML (Stamatakis 2014). Jmodeltest (Darriba *et al.* 2012) indicated that the GTRGAMMA substitution model best fit the data. GTRGAMMA was therefore used in ML

analysis which was repeated 1000 times with 100 bootstrap iterations for each run. The best ML tree was identified using the rapid bootstrapping algorithm in RaXML (Stamatakis 2014) and FigTree (Rambaut and Drummond 2009) was used to visualize the tree using a mid-point root. These reference sequences included published complete mitogenomes from two woolly mammoths (*Mammuthus primigenius*) with GenBank accession numbers EU155210 and EU153449 (Gilbert *et al.* 2008), two African savanna elephants (*Loxodonta africana*) with Genbank accession numbers NC000934 (Hauf *et al.* 2000) and AB443879 (Murata *et al.* 2009), and seven African forest elephants (*Loxodonta cyclotis*) with Genbank accession numbers KY616976, KY616978-9, KJ557423-4, JN673263-4 (Brandt *et al.* 2012a; Finch *et al.* 2014; Meyer *et al.* 2017).

4.6.4 Single nucleotide polymorphisms for nuclear DNA gene regions

To determine the species identity of the elephants from which the tusks had been harvested, we amplified short nDNA regions from three genes (*BGN*, *PHK*, *PLP*) that contain single nucleotide polymorphisms (SNPs) that show fixed character state differences between forest and savanna elephants (Ishida *et al.* 2011) following the amplification procedure described in (Ishida *et al.* 2011; Cui *et al.* 2013) (SI Table 7).

4.6.5 Stable carbon and nitrogen isotope analysis

Elephant tusks grow continuously and incrementally. To average possible seasonal and annual variation, we removed small pieces of ivory extending across multiple growth layers. Sampling attempted to minimise damage to the tusks. In the laboratory, sample surfaces were cleaned by sanding with a Dremel hand drill fitted with an emery disc. Collagen was extracted by demineralising in 0.3M HCl at room temperature for several days to 2 weeks, then rinsed with distilled water to neutrality. Acid was changed every few days. Samples were soaked in 0.1M

NaOH overnight to remove base soluble contaminants and again rinsed with distilled water to neutrality. The samples were then put into pH 3, 0.01M HCl and heated to 70º C for 48 hours to denature the collagen ('gelatinization'), then filtered through 60-90 μm Ezee® filters and lyophilized. Dentine collagen was successfully extracted from 97 of 100 shipwreck ivory samples.

A subset of four tusks was sampled multiple times, targeting successive growth increments to explore variation in isotope values across the lifetime of the elephants. Approximately 10 milligrams of powder were drilled from each increment using a Microdrill. Collagen was extracted from each powder by demineralising in 0.3M HCl at room temperature overnight, then rinsing with distilled water to neutrality, centrifuging between rinses. Due to sample size constraints, only a single $\delta^{13}C$ and $\delta^{15}N$ measurement was obtained for each powder.

Approximately 0.5 milligrams of each extract were weighed into a tin capsule and combusted at 1020ºC in a Thermo Flash Elemental Analyser 2000 coupled to a Delta V Plus mass spectrometer for measurement of carbon and nitrogen isotope ratios as well as elemental compositions (%C, %N). The results are expressed in the delta (δ) notation in parts per thousand (‰), relative to the international standards Vienna Pee Dee Belemnite (VPDB) for carbon and Ambient Inhalable Reservoir (AIR) for nitrogen. All samples other than serial samples were run in duplicate and the values averaged. The standard deviation of repeated measurements (n=40) of homogeneous standard materials was ≤0.2‰ for both $\delta^{13}C$ and $\delta^{15}N$.

4.6.6 GIS data for cartography

We used ArcMap (ESRI 2011, ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute) to create a geographic map of West Africa using the

World Countries (Generalized) layer from ESRI and the UNESCO/UNEP vegetation layers (White 1983). Trading post locations are indicated with current country names.

4.7 ACKNOWLEDGEMENTS

We thank the Government of Namibia, the Namibia National Museum, and Namibia De Beers and their team of consultants and partners for excavating, documenting, and curating the materials from the Oranjemund shipwreck. Special thanks go to archaeologists Bruno Werz and Dieter Noli for their roles in the early stages of the project. We thank the secretariat of the National Heritage Council of Namibia for awarding permit 07/2013 to sample and study the excavated ivory. We thank Esther Goagoses, Henry Nakale, Dawid Kapule, Virimuje Kahuure and Eliot Mowa for assistance with sample collection. We thank colleagues at the University of Pretoria, specifically Catherine Sole, Christian Pirk, Stokana Mahapa and the University of Pretoria DNA sequencing facility (sponsored through an NRF UID78566 facilities grant) for facilitating the laboratory analyses of novel geo-referenced samples from the University of Cape Town collection. Extraction and sequencing of mtDNA from twentieth-century geo-referenced samples was supported by the US Fish and Wildlife Service African Elephant Conservation Fund, grant AFE-1816- F18AP00819. Research support for AdeF was provided through the Francis M. and Harlie M. Clark Research Support Grant (2018), and the Program in Ecology, Evolution and Conservation Biology Research Award, UIUC. AdeF was also supported by the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project number ILLU 875–952 and ILLU-538-939. We thank Madeline Zhu, Ian Newton and John Lanham for help in the stable isotope laboratory. Stable isotope analyses were funded by the South African Research Chairs Initiative of the National Research Foundation and Department of Science and Technology of South Africa (grant no 84407 to JS). AC was

supported by the Claude Leon Foundation and the European Union (FP7- IOF -332165 – TEMBo). Samples were exported under Namibian export heritage permit 07/2013 and CITES permits 166360 and 152037.

4.8 AUTHOR CONTRIBUTION

ADF, AC, JS, SC, RSM and ALR conceived of the study. NMLB curated the shipwreck ivory samples. AC, JS, SC, ADSB, NMLB and RSM provided critical samples and facilities. ADF (genetics) and AC (stable isotopes) conducted experiments and analyses. ADF, AC, JS, SC, ADSB, RSM, and ALR made vital contributions to the interpretation of results. ADF, AC and JS wrote the initial manuscript with major contributions from each of the co-authors. All authors discussed and contributed to the final manuscript.

4.9 FIGURES

4.9.1 Figures

Figure 4.1 The *Bom Jesus* shipwreck had a rich cargo that included several types of gold coins (top image). 10-*cruzado* coins (cross insignia) minted under the reign of King João III of Portugal in 1525, and withdrawn in the 1530s, helped to date the shipwreck (Werz 2010). Other coins included Spanish *excelente* coins showing the profiles of King Ferdinand V of Aragon and Queen Isabella I of Castille (Werz 2010). The shipwreck cargo also included more than 100 unworked elephant tusks of various sizes (bottom image). (Images by Amy Toensing, licensed under the National Geographic Image Collection)

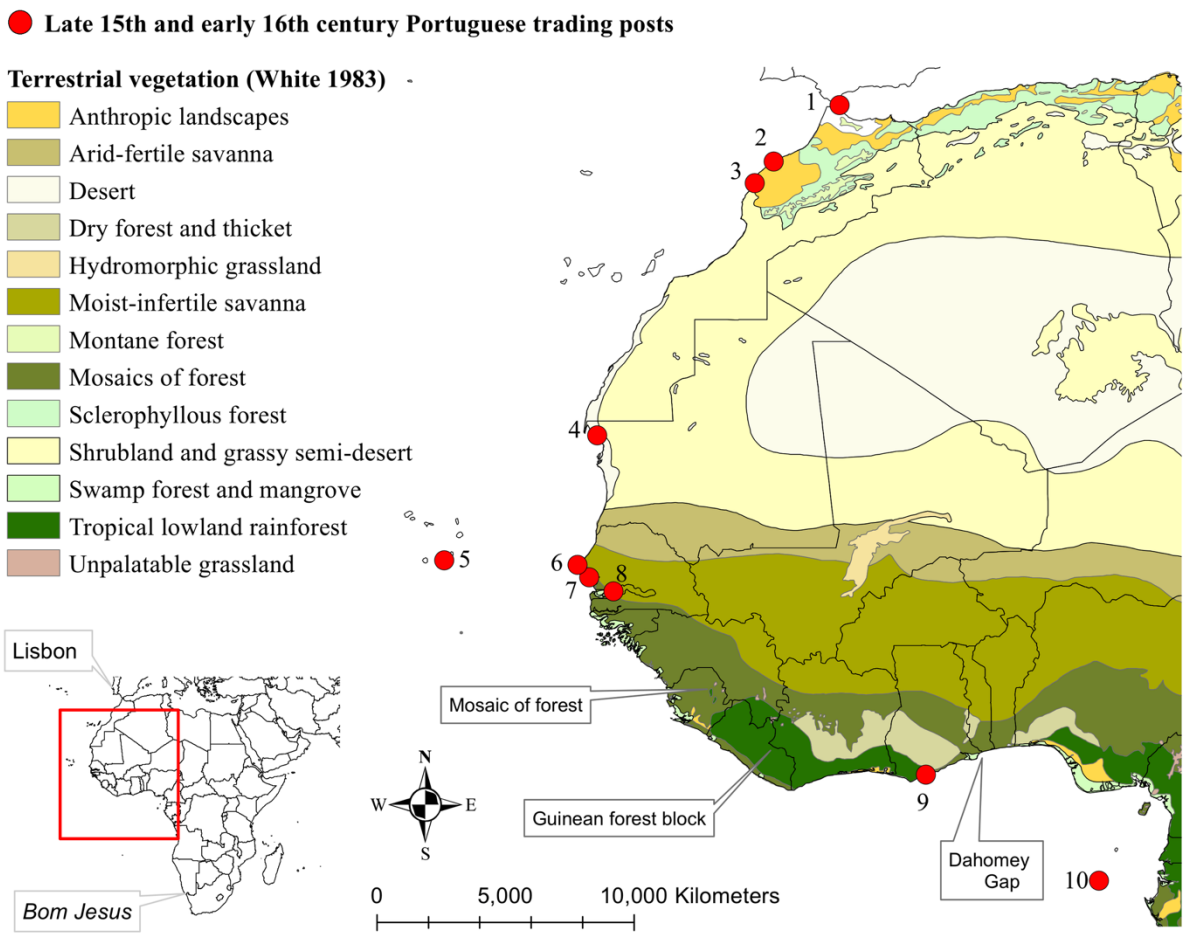

Figure 4.2 West Africa has a variety of terrestrial vegetation types (color-coded), including dense tropical lowland rainforests and mosaics of forests, as well as savannas (White 1983). In the early 16th century, Portuguese merchants operated at several ports and trading regions (red circles) along the West African coast (4). Ports shown are Ceuta (Spain – circle 1), Azemour (Morocco, circle 2), Safi (Morocco, circle 3), Arguim & Cape Blanco (Mauritania, circle 4), Cape Verde (Cape Verde Islands, circle 5), Bezeguiche (Dakar, Senegal, circle 6), Joal (Senegal, circle 7), Sutuco & the Gambia River (The Gambia, circle 8), Elmina (São Jorge de Mina, Ghana, circle 9), and the islands of São Tomé and Príncipe (circle 10). The inset map in the bottom left shows the location of Lisbon in Portugal and of the shipwreck site on the coast of Namibia in southern Africa.

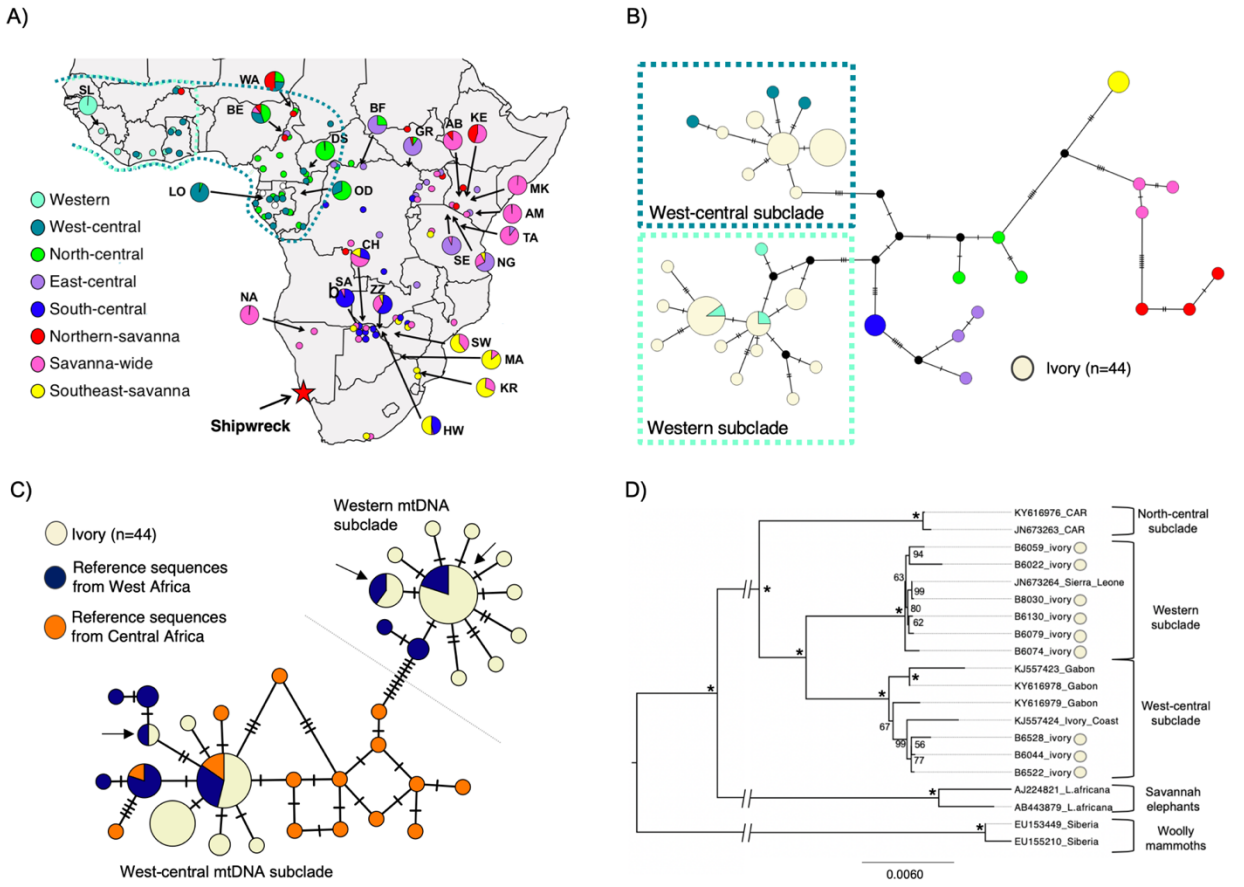

Figure 4.3 The shipwreck ivory originated in West Africa. (A) Ishida et al. (2013) have shown that African elephant mtDNA haplotypes group into 8 well-supported subclades (color-coded) that are geographically structured across Africa (Ishida *et al.* 2013). (B) A median-joining network shows that mtDNA from the shipwreck ivory grouped with only two of the mtDNA subclades reported by Ishida et al. (2013) (Ishida *et al.* 2013). The West-central subclade (dark blue box) is found among elephants in both West and Central Africa, while the Western subclade (light blue box) is found only in West Africa (as is also indicated in panel A by the dashed lines). Note that the ivory haplotypes never grouped within mtDNA subclades that are common among savanna elephants or among central African forest elephants. (C) The ivory mtDNA sequences were also placed in a network with all modern sequences of the Western and West-central mtDNA subclades (haplotypes carried by West African elephants are in dark blue, those from Central Africa are in orange). When the shipwreck ivory shared mtDNA haplotypes with a geographically-referenced sample, the shared haplotype was always reported to be either from West Africa only (black arrows) or from both West and Central Africa, but never from only

Figure 4.3 (cont.) Central Africa. (D) Complete mitogenomes generated for nine of the ancient shipwreck ivory samples grouped with modern elephants from West Africa for both subclades. The mid-point rooted maximum likelihood tree shows bootstrap values (asterisk indicates 100%). Panel A is modified from (Ishida *et al.* 2013) with permission under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by-nc/3.0/legalcode). Panels B and C show pairwise distance-based median joining networks in which branch lengths are proportional to the number of mutational differences (indicated as cross-hatches) between haplotypes, and circle size is related to the number of individuals carrying a haplotype for the ivory samples.

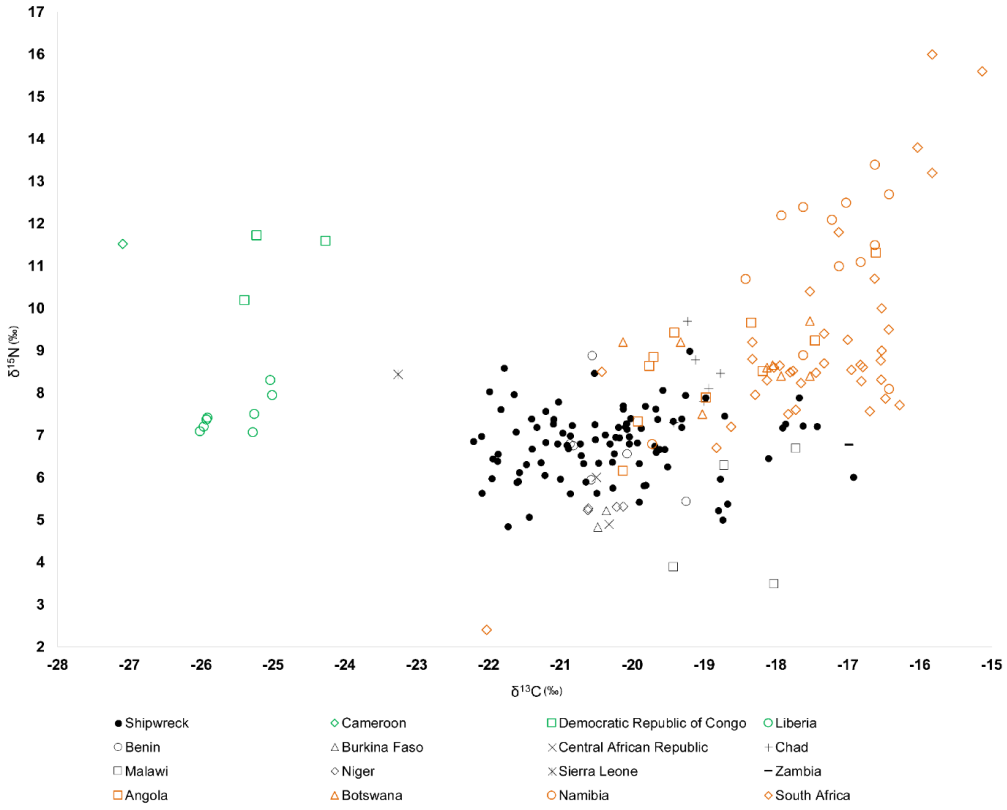

Figure 4.4 Shipwreck ivory $\delta^{13}C$ and $\delta^{15}N$ values, compared with reference samples from elephants from known locations in southern and western Africa (Van der Merwe *et al.* 1990; Codron *et al.* 2012; Coutu *et al.* 2016; Table S7 and S8). Black filled circles are shipwreck ivory samples, other black icons correspond to elephant samples from mesic shrub and wooded savanna environments, orange open icons correspond to elephant samples from drier, open savanna habitats, and green open icons represent elephant samples from closed canopy forest habitats. More positive $\delta^{13}C$ values reflect consumption of substantial quantities of C_4 grass. $\delta^{13}C$ values of reference elephant samples have been corrected for depletion of ^{13}C in atmospheric CO_2 since the Industrial Revolution, due to burning of fossil fuels, to enable direct comparison with the shipwreck archaeological samples (Hellevang and Aagaard 2015). The shipwreck ivory clusters with elephants from environments with mosaic vegetation encompassing both shrub and wooded savanna. The range of values in the shipwreck ivory also overlaps with a variety of elephant habitats, suggesting that the cargo of ivory was sourced from elephants which lived in a variety of habitats with different rainfall and vegetation patterns.

CHAPTER 5: ACCURATE SEX IDENTIFICATION OF ANCIENT ELEPHANT AND OTHER ANIMAL REMAINS USING LOW-COVERAGE DNA SHOTGUN SEQUENCING DATA[3]

5.1 ABSTRACT

Sex identification of ancient animal biological remains can benefit our understanding of historical population structure, demography and social behavior. Traditional methods for sex identification (e.g. osteological and morphometric comparisons) may be ineffective when animal remains are not well preserved, when sex distinguishing characteristics have not yet developed, or where organisms do not exhibit sex-associated phenotypic dimorphisms. Here we adapt a method developed for human sex determination so that it can be used to identify the sex of ancient and modern animal taxa. The method identifies sex by calculating the ratio of DNA reads aligning to the X chromosome to DNA reads aligning to autosomes (termed the Rx ratio). We tested the accuracy of this method using low coverage genomes from 15 modern elephants for which sex was known. We then applied this method to ancient elephant ivory samples for which sex was unknown, and describe how this method can be further adapted to the genomes of other taxa. This method may be especially useful when only low-coverage genomic data is obtainable. Furthermore, because this method relies on only the *X* and not the *Y* chromosome, it can be used to determine the sex of organisms for which a reference genome was obtained from a female or for which only the *X* chromosome is reported. Such taxa include the domestic cat, sheep, goat, and horse; and non-domesticated animals such as the Sumatran orangutan, western lowland gorilla and meerkat.

[3] Chapter 5 has previously been published as de Flamingh, A., A. Coutu, A. L. Roca, and R. S. Malhi, 2020 Accurate Sex Identification of Ancient Elephant and Other Animal Remains Using Low-Coverage DNA Shotgun Sequencing Data. G3 Genes|Genomes|Genetics DOI: 10.1534/g3.119.400833

5.2 INTRODUCTION

Identifying the sex of animals can yield insights into population structure (Schoener and Schoener 1980; Bodkin *et al.* 2000), demographic histories (Sæther *et al.* 2003; Heyer *et al.* 2012) and social interactions (Pedersen *et al.* 1990; Lonsdorf *et al.* 2014). It can add to knowledge of extinct and extant animal populations and reveal how they have changed across time. Sex identification can aid our understanding of extinct animal biology (Allentoft *et al.* 2010), past hunting practices and domestication (Collier and White 1976; d'Errico and Vanhaeren 2002). For many ancient or historical samples, however, the sex of specimens is unknown. Sex identification may be hindered when remains are very degraded or only partially preserved, when remains are from young individuals where sex distinguishing characteristics have yet to develop, or when remains are from taxa that do not exhibit phenotypic sexual dimorphism (Hamilton *et al.* 1986). Such factors may preclude sex identification through traditional methods such as osteological or morphometric comparison (measurements of skeletal ratios/aspects) (Safont *et al.* 2000; Rogers 2005; Bruzek and Murail 2006).

Molecular sex identification circumvents these issues, requiring only a small sample for DNA analysis. For ancient samples with a low quantity and quality of DNA (Quincey *et al.* 2013), molecular methods test for DNA authenticity by determining whether amplified DNA exhibits damage patterns typical of ancient DNA (Jónsson *et al.* 2013). Molecular methods therefore permit sex identification of degraded or partial specimens, from young and from sexually monomorphic taxa. Molecular sex identification methods have involved the analysis of genes associated with male and female sex chromosomes in birds (Griffiths *et al.* 1998; Fridolfsson and Ellegren 1999), reptiles (Quinn *et al.* 2009), mammals (Sullivan *et al.* 1993; Gibbon *et al.* 2009), and fish (Chen *et al.* 2007). For example, in some mammals, molecular sex

identification involves differentiating between amelogenin gametologues on the *X* and *Y* chromosomes (Sullivan *et al.* 1993; Gibbon *et al.* 2009).

For sex identification of archeological human remains, Mittnik *et al*. (2016) developed a method that uses low coverage whole genome data to calculate the Rx ratio, which compares DNA sequence reads that align to the *X* chromosome to DNA sequence reads that align to autosomal chromosomes. The Rx ratio is different for females and males, since they have two or one X-chromosomes, respectively. The Rx ratio would be expected to be ca. 1.0 for females and 0.5 for males. Mittnik *et al*. (2016) identified individuals as female if the Rx 95% CI lower bound was higher than 0.80, and as male if the 95% confidence interval (CI) upper bound for Rx was lower than 0.60.

Here, we present an extension and expansion of the method of Mittnik *et al*. (2016) to permit sex identification of ancient and modern samples of non-human taxa. We adjust the Rx equation to mathematically account for different chromosome numbers across animal taxa. Our method, in principle, allows for accurate sex determination of any organism with *XY* sex determination for which a reference genome is available with chromosome-level resolution. We verify the method using low-coverage genomes from 15 modern elephants for which sex is known, and apply this method successfully to low-coverage genomes of ten ancient elephant ivory samples for which sex was unknown. These ancient ivory samples are from a 16th Century shipwreck uncovered in Namibia and believed to be the *Bom Jesus*, a Portuguese trading ship lost in 1533 en route to India (Werz 2010; Alves 2011).

5.3 MATERIALS AND METHODS

5.3.1 DNA extraction and shotgun sequencing

DNA was extracted from skin biopsy samples from 15 African elephants for which sex was recorded in the field when the samples were collected. The modern elephant samples were from nine females and six males (Table 5.1). Genomic libraries were constructed for the 15 modern elephants at the UIUC Core Sequencing Facility using TruSeq® DNA library preparation. To generate low-coverage genomes for the modern elephant samples, we sequenced (150bp paired-end) the 15 samples as part of a larger pool of samples in a single HiSeq 4000 lane.

DNA from the ancient ivory was extracted following methods described in Cui *et al.* (2013). Cui *et al.* (2013) provide details, for example, on starting template amounts (0.20g per ancient sample) and treatment protocols. Ancient DNA work (extractions and genomic library preparation) was conducted in the Malhi Ancient DNA Laboratory, which is dedicated exclusively to studies involving ancient DNA, at the Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign (UIUC). All rounds of DNA extraction included a negative control to verify that reagents and equipment were not contaminated and that there was no cross-contamination between samples, and not more than eight samples were processed at any one time. Libraries for the ten ancient ivory samples were constructed using the NEBNext® Ultra II^{TM} DNA Library Prep kit and NEBNext® Multiplex Oligos (Unique Dual Indexes) for Illumina®. Because ancient DNA are prone to have cytosine to uracil nucleotide base changes (Hofreiter *et al.* 2001), the extracted ancient DNA was pre-treated with USER (Uracil-Specific Excision Reagent) enzyme. The modern and ancient libraries were pooled

separately, and each pool was shotgun sequenced on a HiSeq 4000 platform at the UIUC Core Sequencing Facility.

5.3.2 Bioinformatic analyses and Rx based sex identification

Sample reads were de-multiplexed and trimmed using the program FastP v.0.19.6 (Chen *et al.* 2018) to have a minimum sequence length of 25bp. Reads were aligned to the chromosome-level assembly of the African savanna elephant genome (*Loxodonta africana* assembly Loxafr4.0, Broad Institute (Palkopoulou *et al.* 2018)) using bowtie2 (Langmead and Salzberg 2012) with the local alignment option, and capping fragment length at 1000bp. Aligned sequences were transformed to BAM format in SAMtools v. 1.1 (Li *et al.* 2009). Using SAMtools, BAM files were filtered to remove unmapped reads and reads with a quality score less than 30, then sorted and indexed, with PCR duplicates marked and removed with the Picard Toolkit v. 2.10.1 ("Picard Toolkit." 2019. Broad Institute). Index statistics for BAM files were generated using "idxstats" in SAMtools (Li *et al.* 2009).

The Rx_identifier.r script of Mittnik *et al*. (2016) was modified to accommodate the number of chromosome pairs found in elephants, which is different from the number in humans, for which the script was originally developed (see Appendix D for a stepwise protocol of how to modify this script for any organism that has a chromosome-level reference genome and *XY* sex determinism). We verified that the row numbers in the Rx_identifier script corresponded to the correct chromosome identities in our sorted idxstat files. The modified Rx_identifier.r script was then implemented using the program R v. 3.3.3 (R-Development-Core-Team 2017) and the indxstat files as input. Output statistics for each sample included the Rx ratio, and sex identification based on the data ranges of Mittnik *et al*. (2016), where a sample was identified as male if its 95% confidence interval (CI) upper bound for Rx was lower than 0.60 and identified

as female if its Rx 95% CI lower bound was higher than 0.80. The 95% CI was computed as Rx±1.96SE (standard error), where the SE measures the amount of variability in the Rx mean compared to autosomes (22 for humans, 27 for elephants). We determined whether sequence coverage was sufficient by performing a linear regression of the number of sequenced and mapped reads on each chromosome against the number of reference reads. Output statistics were visualized by plotting individual Rx ratios (Figure 5.1) using R v. 3.3.3 (R-Development-Core-Team 2017). The bioinformatic analyses were repeated using BWA (Li and Durbin 2010) to check for inconsistencies that could be associated with sequence aligner choice, but no inconsistencies were observed and sex identification was completely consistent between the two analyses. Ancient DNA damage patterns were verified by aligning trimmed reads to the African savannah elephant genome (LoxAfr 4.0) using BWA (Li and Durbin 2010) and quantifying damage in mapDamage2 (Jónsson *et al.* 2013) using a fragment size of 70bp.

To determine how effective the Rx method is for determining the sex of samples with even lower coverages that ours, we subsampled the existing ancient ivory data to include datasets of approximately 10 000 and 1 000 reads. We used Sambamba (Tarasov et al. 2015) to subsample datasets, and reanalyzed the subsampled datasets using the Rx method.

5.3.3. Data availability

All idxstats files for modern and ancient elephant genomes, and all Rx ratio result files for the 10 000- and 1 000-read subsampled files are available from GSA Journals figshare portal (https://gsajournals.figshare.com/s/cf0f4f71a5c88d87361c). The most recent update of the savanna elephant reference genome (*LoxAfr4*) is available at ftp://ftp.broadinstitute.org/pub/assemblies/mammals/elephant/loxAfr4/. R-scripts and a step-by-step description of how to adapt the R-script to any species chromosome-level genome can be

found at https://github.com/adeflamingh/de_Flamingh_et_al_2020_G3.git or as Appendix D (step-by-step protocol), or as part of the supplementary material on GSA Journals figshare (https://gsajournals.figshare.com/s/cf0f4f71a5c88d87361c). The study was conducted under the University of Illinois Institutional Animal Care and Use Committee approved protocol number 18042. Samples were imported through a CITES permit.

5.4 RESULTS

Although all ancient samples were pre-treated with USER enzyme, which may potentially mask damage patterns (see Methods), DNA damage patterns in the ancient ivory were still evident and typical of ancient DNA (Figure E.1). Each of the ancient samples showed increased rates of C to T and G to A mismatches relative to the reference genome, as would be expected in authentic ancient DNA.

The adapted Rx_identifier.R script (available from https://github.com/adeflamingh/de_Flamingh_et_al_2020_G3.git) was able to identify sex to all 15 modern individuals (nine females, six males) with 100% accuracy (Figure 5.1; Table 5.1). For the ancient DNA remains, the adapted script identified eight individuals as male, and two individuals as female (Figure 5.1; Table 5.2). Linear regressions of the number of reference genome reads with the number of mapped reads resulted in significant F-statistic values ($p<0.001$) for both modern and ancient remains, indicating that the sequence coverage for all genomes was sufficient for accurate sex determination.

The Rx ratio method effectively identified the sex when using datafiles with > 100 000 reads (all 95% CI are within the specified Rx cut-off values; Table E.2). The method was mostly effective when using subsampled ancient ivory datafiles with ~10 000 reads (only sample B6079 had 95% CI outside of Rx cut-off values), but proved less effective when using subsampled

datasets with ~1000 reads (the span of the 95% CI increased for all samples and four samples had 95% CI outside of Rx cut-off values; Figure E.2).

5.5 DISCUSSION

We adapted a method previously developed (Mittnik *et al.* 2016) for sex identification of human remains for use with non-human taxa, and successfully identified the sex of modern and ancient elephants from low coverage genome data. Because the Rx ratio sex identification method presented in this study relies only on the *X* and not the *Y* chromosome, it can be used to identify the sex of organisms in which the reference genome was obtained from a female animal or where only the *X* and not the *Y* chromosome is reported in the reference genome assembly for the taxon. Such taxa would include (but not be limited to) the domestic cat, sheep, goat, horse, dromedary camel, European rabbit; and also include many wild animals such as the Sumatran orangutan, western lowland gorilla, gelada and meerkat (Table E.1). Being able to identify the sex of samples could benefit agricultural studies on domesticated animals, and could inform conservation initiatives that focus on non-domestic wildlife. Because this method is amenable to low coverage data from low quantity DNA (e.g. ancient or degraded DNA), it can be employed as a non-invasive approach to identifying sex of endangered or rare species, for example, through the analysis of DNA from hair tufts (McKelvey *et al.* 2006; Stanton *et al.* 2016) or herbivore scat (Huber *et al.* 2002). By requiring only minute quantities of DNA as a starting template, the method could be extended to other types of degraded DNA such as archival samples from museum collections (Wandeler *et al.* 2003; Bi *et al.* 2013) or forensic samples (Jobling and Gill 2004; Alaeddini *et al.* 2010).

Sex identification using the Rx ratio could be adapted to any taxa that exhibit *XY* sex determination for which a chromosome-level genome assembly is available. It should be possible

to further extend the method to taxa that have a *ZW* sex determination system, in which males are the homogametic sex *ZZ*, and females have *Z* and *W* chromosomes. Such taxa include birds (Chue and Smith 2011), amphibians (Nakamura 2009) and crustaceans (Cui *et al.* 2015). For *ZW* sex determination systems, individuals should be identified as male (*ZZ*) if the lower bound of their 95% Rx ratio CI is approximately 0.8 or higher, and female (*ZW*) if the upper bound of the 95% Rx ratio CI is approximately 0.6 or less. For *ZW* sex determination the script should be adapted so that the *Z* chromosome replaces the *X* chromosome in the Rx_identifier.R script, and the *W* chromosome replaces the *Y* chromosome (if it is present in the reference genome). Again, since the script could rely only on the *Z* chromosome and not the *W* chromosome, this method may be used on any individual, male or female, with *ZW* sex determination if there is a chromosome-level reference genome assembly available for that species. Future studies would be needed to validate the use of the adapted script on animals other than elephants or humans.

We investigated whether the Rx method can effectively identify the sex of individuals when using genome coverage even lower than that of the ancient ivory samples. We found that there is a substantial broadening of the 95% CI as the read count of the datafile decreases (Figure E.2). We suggest that the Rx cut-off values presented by Mittnik et al 2016, and in this paper, may be useful indicators of the ability of Rx script to accurately and precisely identify individual sex, and caution users to be less confident in sex identification if the confidence intervals extend beyond these cut-off values.

The Rx ratio method was successfully used here on low coverage genomic data from both modern and ancient (Table E.2) elephants. The ability to accurately identify sex based on low coverage data may be especially useful with ancient samples with DNA of low quantity and quality (Quincey *et al.* 2013), and for studies that index and pool a large number of individuals

for sequencing (e.g. PoolSeq studies). Such studies may have low coverage per individual, but many individuals may be indexed and pooled to represent a population. The limited requisites and ease of adaptation and implementation of this method would allow for convenient and effective identification of the sex of modern and ancient animal remains.

5.6 ACKNOWLEDGEMENTS

This project was supported by US Fish and Wildlife Service African Elephant Conservation Grant AFE-1816-F18AS00055, and by a Program in Ecology, Evolution and Conservation Biology Research Award to AdF through the Cooperative State Research, Education, and Extension Service, US Department of Agriculture, under project numbers ILLU 875–952 and ILLU-538-939. We thank the CORE sequencing facility for their help with constructing the genomic DNA TruSeq libraries. For technical and other assistance, we thank T. Perrin-Stowe and Y. Ishida. We thank Judith Sealy and Shadreck Chirikure at the University of Cape Town, South Africa, and David Reich at Harvard Medical School, Boston, USA, for providing access to the ancient ivory samples and modern elephant genomic data.

5.7 AUTHOR CONTRIBUTION STATEMENT

ADF, ALR and RSM conceptualized the project. ADF and AC conducted molecular laboratory work. ADF conducted the bioinformatic analysis. ADF, ALR and RSM wrote the main manuscript text. All authors contributed by editing and reviewing the manuscript prior to publication.

5.8 TABLES AND FIGURE

5.8.1 Tables

Table 5.1 Known sex of modern elephants, and predicted sex using the Rx ratio.

Sample ID	Rx ratio[1]	95% Confidence Intervals (CI)[2]	Known sex	Predicted sex
DS1531	0.9348111	0.9241519 0.9454702	Female	Female
DS1548	0.4844121	0.4781392 0.4906849	Male	Male
DS1514	0.9300965	0.9194486 0.9407445	Female	Female
DS1543	0.4918646	0.4866054 0.4971237	Male	Male
DS1506	0.4878141	0.4826157 0.4930124	Male	Male
LO3503	0.9321284	0.9215853 0.9426715	Female	Female
LO3509	0.9415045	0.9308473 0.9521617	Female	Female
LO3511	0.9287396	0.917744 0.9397352	Female	Female
LO3521	0.8712463	0.8574852 0.8850075	Female	Female
LO3514	0.8810152	0.8687895 0.893241	Female	Female
GR0041	0.91914	0.9082285 0.9300516	Female	Female
GR0037	0.4560117	0.4489408 0.4630825	Male	Male
WA4013	1.015361	1.001641 1.029081	Female	Female
WA4020	0.496401	0.491036 0.5017661	Male	Male
BE4059	0.5012983	0.4956742 0.5069223	Male	Male

[1]The Rx ratio compares DNA sequence reads that align to the *X* chromosome to DNA sequence reads that align to autosomal chromosomes, and would be expected to be ca. 1.0 for females and 0.5 for males. [2] The top value represents lower bound of the 95% CI and the lower value represents the upper bound of the 95% CI.

Table 5.2 Predicted sex of ancient elephant samples using the Rx ratio

Ancient sample ID	Read count	Rx ratio[1]	95% Confidence intervals (CI)[2]	Predicted sex
B6025	991,719	0.4917869	0.4827309 0.5008429	Male
B8030	6,186,805	1.00599	0.9925127 1.019467	Female
B6059	3,672,375	0.4827319	0.4723486 0.4931151	Male
B6074	864,649	0.4912222	0.4844642 0.4979802	Male
B6022	1,848,409	0.4955726	0.4894958 0.5016494	Male
B6062	919,574	0.9684064	0.9464218 0.9903909	Female
B6051	460,668	0.4908337	0.476899 0.5047683	Male
B6130	17,797,452	0.4745888	0.4611754 0.4880023	Male
B6520	141,855	0.4915878	0.4808738 0.5023018	Male
BB6079	5,729,007	0.477581	0.4637086 0.4914535	Male

[1]The Rx ratio compares DNA sequence reads that align to the *X* chromosome to DNA sequence reads that align to autosomal chromosomes, and would be expected to be ca. 1.0 for females and 0.5 for males. [2] The top value represents lower bound of the 95% CI and the lower value represents the upper bound of the 95% CI.

5.8.2 Figure

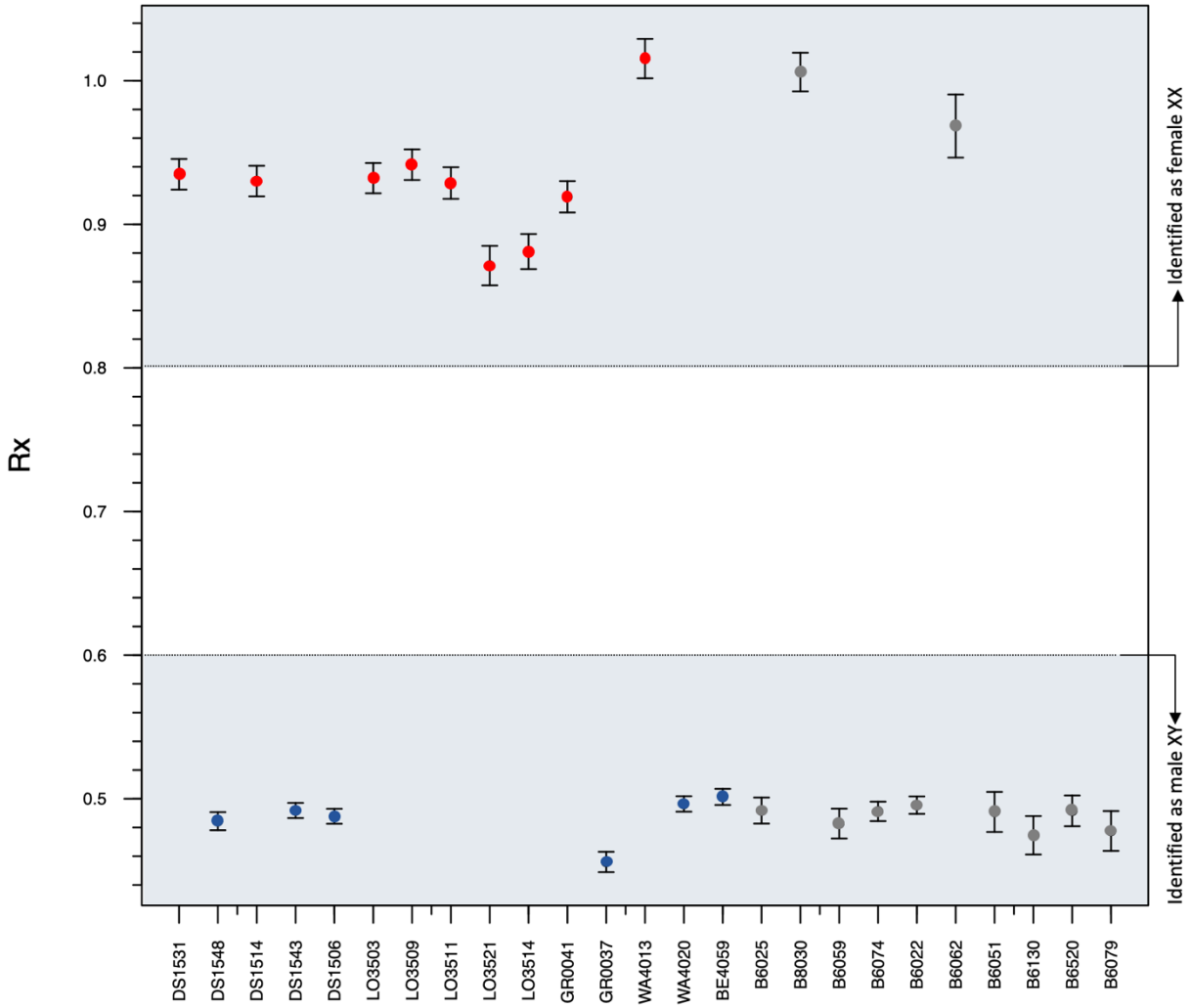

Figure 5.1 Rx values for modern elephants of known sex (red and blue) and ancient elephant samples of unknown sex (gray). Rx is the ratio of sequence read alignments to the *X* chromosome compared to sequence read alignments to all autosomes. Rx values are shown for low coverage modern elephant genomes of known males (blue) and females (red), and for ancient elephant genomes for ivory samples of previously unknown sex (gray). An Rx ratio with an upper 95% CI of less than 0.6 indicates male sex, and an Rx ratio with a lower 95% CI that is greater than 0.8 indicates female sex.

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSIONS

Elephant population numbers are rapidly diminishing as a consequence of habitat loss and poaching for ivory. Pro-active, preventative and restorative conservation efforts are critical. Conservation scientists and stakeholders could, for example, focus on the restoration and maintenance of distributional ranges to link formerly isolated populations. Such initiatives can be considered functional if they disrupt population isolation by allowing dispersal or migration between nearby populations. Linking isolated populations would have beneficial genetic consequences if spatial isolation has been a driving force of genetic erosion (e.g., population bottlenecks and the subsequent loss of genetic diversity due to drift).

In Chapter 2 I developed and evaluated an approach to identify and delineate functional landscape linkages (FLL) for African elephant conservation planning. I identified FLL between African savannah elephant populations that occur across 7 southern African countries by combining habitat suitability modelling with gene flow patterns. I found a pronounced nonlinear response to habitat suitability which suggests that elephant movement and gene flow are mostly impacted by very unsuitable habitats, and that moderately unsuitable habitat impedes connectivity between elephant populations to a lesser degree. It may therefore be beneficial for future studies to consider both behavior and space use when delineating FLL for African elephants.

The results of Chapter 3 in this thesis provide support for conservation initiatives that aim at increasing connectivity through FLL. Chapter 3 shows that functional entities of connected elephant populations may prevent or counteract genetic consequences of isolation (e.g., decreased genetic diversity associated with population isolation). Kruger National Park's elephant population did not show evidence of genetic isolation, a depleted gene pool or a strong

founder effect. This is likely due to historical founding events involving more than one source population, along with a rapid increase to a large population size, and possible subsequent immigration into Kruger that augmented the gene pool. Chapter 3 highlights the importance of migration and gene flow for maintaining genetic diversity in populations. The results of Chapter 3 suggest that Kruger forms part of a functional entity in which migration helped to maintain a relatively diverse gene pool, and this Chapter therefore emphasizes the need for conservation initiatives such as transfrontier conservation areas that aim at maintaining connectivity between populations. Initiatives that maintain connectivity may provide a sustainable, self-regulating management approach for elephants in southern Africa that simultaneously upholds genetic diversity and gene flow.

The outcomes of Chapter 4 benefit elephant conservation by increasing our understanding of historical elephant population ecology and genetics. Chapter 4 shows that contemporary West African forest elephant populations may have limited genetic diversity compared to the genetic diversity found in historical populations. Only two of ten Western subclade mtDNA haplotypes present among the 16th century tusks have been reported among contemporary populations. The decreased genetic diversity observed in contemporary populations is perhaps not surprising given that elephant distributions in West Africa have been reduced by 93% in the last century (Roth and Douglas-Hamilton 1991). In addition, forest elephant populations have been targeted by ivory poachers (Alers *et al.* 1992; Wasser *et al.* 2015) and in some areas the number of forest elephants have declined by as much as two-thirds (Alers *et al.* 1992). Decreased genetic diversity is of conservation concern because it has been associated with negative demographic and physiological outcomes, such as expression of deleterious alleles, reduced reproductive fitness and increased risk of population extirpation

(Allendorf *et al.* 2013). The overall decrease in forest elephant numbers may also have drastic ecological implications since forest elephants are known to be the primary dispersers of many Afrotropical forest tree species (Beaune *et al.* 2013; Poulsen *et al.* 2018). The outcomes of Chapter 4 therefore highlight the need for pro-active and preventative conservation strategies that could conserve the remaining forest elephant populations, and so doing prevent any further loss of genetic diversity.

In Chapter 4 of this thesis I adapted a method previously developed (Mittnik *et al.* 2016) for sex identification of human remains for use with non-human taxa, and successfully identified the sex of modern and ancient elephants from low coverage genome data. Because the Rx ratio sex identification method presented in this study relies only on the *X* and not the *Y* chromosome, it can be used to identify the sex of organisms in which the reference genome was obtained from a female animal or where only the *X* and not the *Y* chromosome is reported in the reference genome assembly for the taxon. Such taxa would include (but not be limited to) the domestic cat, sheep, goat, horse, dromedary camel, European rabbit; and also include many wild animals such as the Sumatran orangutan, western lowland gorilla, gelada and meerkat. Sex identification of ancient animal biological remains can benefit conservation by increasing our understanding of historical population structure, demography and social behavior.

This thesis provides a toolset and framework for future conservation genetic studies that focus on African elephant populations. However, additional research is needed to quantify the effects of genetic isolation on African elephant population demography and phenotypic traits. To this end, researchers need to identify and characterize the current genetic diversity present in African elephant populations. As demonstrated in this thesis, information on the genetic characteristics of extant and extinct elephant populations can benefit elephant conservation by

allowing conservation stakeholders to identify populations and areas in the landscape that should be prioritized for conservation.

Researchers and conservation stakeholders, however, should also consider factors other than those investigated in this thesis when determining which populations and areas to conserve. For example, factors such as land ownership (Pinter-Wollman 2012), human-elephant conflict mitigation (Jackson *et al.* 2008; Pinter-Wollman 2012), the direct and indirect impact of protected area creation on local indigenous communities (Baldus *et al.* 2007), protected area overlap with poaching hotspots (Osborn and Parker 2003; Booth and Dunham 2016), and other sociopolitical factors need to be integrated into conservation decisions. Strategies that consider all of these factors could allow for the development of well-thought-out initiatives for the successful conservation of elephant populations in Africa.

REFERENCES

van Aarde, R. J., and T. P. Jackson, 2007 Megaparks for metapopulations: addressing the causes of locally high elephant numbers in southern Africa. Biol. Conserv. 134: 289–297.

van Aarde, R., I. Whyte, and S. Pimm, 1999 Culling and the dynamics of the Kruger National Park African elephant population, pp. 287–294 in *Animal Conservation forum*, Cambridge University Press.

Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos, 2008 Estimating space-use and habitat preference from wildlife telemetry data. Ecography (Cop.). 31: 140–160.

Academia Das Ciências De Lisboa, 1979 *Memoria Das Armadas Que De Portugal Passaram à India*. Lisboa: Academia Das Ciências.

Afonso, L. U., and J. da Silva Horta, 2013 Afro-Portuguese Olifants with hunting scenes (c. 1490-c. 1540). Mand. Stud. 79–97.

Alaeddini, R., S. J. Walsh, and A. Abbas, 2010 Forensic implications of genetic analyses from degraded DNA—a review. Forensic Sci. Int. Genet. 4: 148–157.

de Alencastro, L. F., 2007 The Economic Network of Portugal's Atlantic World', Francisco Bethencourt en Diogo Ramada Curto. Port. Ocean. Expans. 1400-1800 109–137.

Alers, M. P. T., A. Blom, C. S. Kiyengo, T. Masunda, and R. F. W. Barnes, 1992 Preliminary assessment of the status of the forest elephant in Zaire. Afr. J. Ecol. 30: 279–291.

Allendorf, F. W., and G. Luikart, 2009 *Conservation and the genetics of populations*. John Wiley & Sons.

Allendorf, F. W., G. Luikart, and S. N. Aitken, 2013 *Conservation and the genetics of populations*. Blackwell Publishing, Wiley and Sons, West Sussex, UK.

Allentoft, M. E., M. Bunce, R. P. Scofield, M. L. Hale, and R. N. Holdaway, 2010 Highly skewed sex ratios and biased fossil deposition of moa: ancient DNA provides new insight on New Zealand's extinct megafauna. Quat. Sci. Rev. 29: 753–762.

Alves, F. J. S., 2011 The 16th century Portuguese shipwreck of Oranjemund, Namibia. Rep. Mission. Carried out by Port. Team 2008 2009.

Ambrose, S. H., 1990 Preparation and characterization of bone and tooth collagen for isotopic analysis. J. Archaeol. Sci. 17: 431–451.

Amundson, R., A. T. Austin, E. A. G. Schuur, K. Yoo, V. Matzek *et al.*, 2003 Global patterns of the isotopic composition of soil and plant nitrogen. Global Biogeochem. Cycles 17.

Andréasson, H., U. Gyllensten, and M. Allen, 2002 Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. Biotechniques 33: 402–411.

Anthony, B., and G. Avery, 2008 *Elephant management: A scientific assessment for South Africa*. NYU Press.

Aranibar, J. N., L. Otter, S. A. Macko, C. J. W. Feral, H. E. Epstein *et al.*, 2004 Nitrogen cycling in the soil–plant system along a precipitation gradient in the Kalahari sands. Glob. Chang. Biol. 10: 359–373.

Archie, E. A., J. A. Hollister-Smith, J. H. Poole, P. C. Lee, C. J. Moss *et al.*, 2007 Behavioural inbreeding avoidance in wild African elephants. Mol. Ecol. 16: 4138–4148.

Archie, E. A., J. E. Maldonado, J. A. HOLLISTER-SMITH, J. H. Poole, C. J. Moss *et al.*, 2008 Fine-scale population genetic structure in a fission–fusion society. Mol. Ecol. 17: 2666–2679.

Baldus, R. D., R. Hahn, C. Ellis, and S. D. DeLeon, 2007 Connecting the world's largest elephant ranges: The Selous-Niassa corridor. Peace Park. Conserv. Confl. Resolut. 109–126.

Barnes, R. F. W., K. L. Barnes, M. P. T. Alers, and A. Blom, 1991 Man determines the distribution of elephants in the rain forests of northeastern Gabon. Afr. J. Ecol. 29: 54–63.

Barton, K., 2009 MuMIn: multi-model inference. R package version 1. 0. 0. http://r-forge. r-project. org/projects/mumin/.

Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen *et al.*, 2015 Package 'lme4.' Convergence 12.

Beaumont, M., and J. Wang, 2019 Conservation genetics. Handb. Stat. Genomics Two Vol. Set 440–457.

Beaune, D., B. Fruth, L. Bollache, G. Hohmann, and F. Bretagnolle, 2013 Doom of the elephant-dependent trees in a Congo tropical forest. For. Ecol. Manage. 295: 109–117.

Becker, R. A., J. M. Chambers, and A. R. Wilks, 1988 The New S Language. Wadsworth & Brooks/Cole. Comput. Sci. Ser. Pacific Grove, CA.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen *et al.*, 2013 Unlocking the vault: next-generation museum population genomics. Mol. Ecol. 22: 6018–6032.

Blake, J. W., 1942 Europeans in West Africa 1450-1560, printed for the Hakluyt Society, Vol. 1. I (London, 1942).

Blanc, J., 2008 *Loxodonta africana*. The IUCN Red List of Threatened Species 2008: e. T12392A3339343.

Bodkin, J. L., A. M. Burdin, and D. A. Ryazanov, 2000 Age-and sex-specific mortality and population structure in sea otters. Mar. Mammal Sci. 16: 201–219.

Bonafini, M., M. Pellegrini, P. Ditchfield, and A. M. Pollard, 2013 Investigation of the 'canopy effect'in the isotope ecology of temperate woodlands. J. Archaeol. Sci. 40: 3926–3935.

Booth, V. R., and K. M. Dunham, 2016 Elephant poaching in Niassa Reserve, Mozambique: population impact revealed by combined survey trends for live elephants and carcasses. Oryx 50: 94–103.

Bouckaert, R., and J. Heled, 2014 DensiTree 2: Seeing trees through the forest. BioRxiv 12401.

Boxer, C. R., 1991 *The Portuguese seaborne empire, 1415-1825*. Exeter: Carcanet.

Braaker, S., U. Kormann, F. Bontadina, and M. K. Obrist, 2017 Prediction of genetic connectivity in urban ecosystems by combining detailed movement data, genetic data and multi-path modelling. Landsc. Urban Plan. 160: 107–114.

Brandt, A. L., Y. Ishida, N. J. Georgiadis, and A. L. Roca, 2012a Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. Mol. Ecol. 21: 1175–1189.

Brandt, A. L., Y. Ishida, N. J. Georgiadis, and A. L. Roca, 2012b Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. Mol. Ecol. 21: 1175–1189.

Briggs, A. W., U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso *et al.*, 2007 Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. 104: 14616–14621.

Bright, E. A., A. N. Rose, M. L. Urban, and J. J. McKee, 2017 LandScan 2016.

Bruzek, J., and P. Murail, 2006 Methodology and reliability of sex determination from the skeleton, pp. 225–242 in *Forensic anthropology and medicine*, Springer.

Burkart, S., F. Gugerli, J. Senn, R. Kuehn, and J. Bolliger, 2016 Evaluating the functionality of expert-assessed wildlife corridors with genetic data from roe deer. Basic Appl. Ecol. 17: 52–60.

Burnham, K. P., and D. R. Anderson, 2004 Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33: 261–304.

Cameron, A. C., R. B. Page, J. I. Watling, C.-A. M. Hickerson, and C. D. Anthony, 2019 Using a comparative approach to investigate the relationship between landscape and genetic connectivity among woodland salamander populations. Conserv. Genet. 20: 1265–1280.

de Castro, F. V., 2005 *The pepper wreck: a Portuguese Indiaman at the mouth of the Tagus river*. Texas A&M University Press.

Castro, F., N. Fonseca, and A. Wells, 2010 Outfitting the pepper wreck. Hist. Archaeol. 44: 14–34.

Cerling, T. E., J. A. Hart, and T. B. Hart, 2004 Stable isotope ecology in the Ituri Forest. Oecologia 138: 5–12.

Cerling, T. E., P. Omondi, and A. N. Macharia, 2007 Diets of Kenyan elephants from stable isotopes and the origin of confiscated ivory in Kenya. Afr. J. Ecol. 45: 614–623.

Cerling, T. E., G. Wittemyer, J. R. Ehleringer, C. H. Remien, and I. Douglas-Hamilton, 2009 History of animals using isotope records (HAIR): a 6-year dietary history of one family of African elephants. Proc. Natl. Acad. Sci. 106: 8093–8100.

Chamaillé-Jammes, S., M. Valeix, and H. Fritz, 2007 Managing heterogeneity in elephant distribution: interactions between elephant population density and surface-water availability. J. Appl. Ecol. 44: 625–633.

Chen, S.-L., J. Li, S.-P. Deng, Y.-S. Tian, Q.-Y. Wang *et al.*, 2007 Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar. Biotechnol. 9: 273–280.

Chen, S., Y. Zhou, Y. Chen, and J. Gu, 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884–i890.

Chirikure, S., A. Sinamai, E. Goagoses, M. Mubusisi, and W. Ndoro, 2010 Maritime archaeology and trans-oceanic trade: a case study of the Oranjemund shipwreck cargo, Namibia. J. Marit. Archaeol. 5: 37–55.

Chue, J., and C. A. Smith, 2011 Sex determination and sexual differentiation in the avian model. FEBS J. 278: 1027–1034.

CILSS - Comité Permanent Inter-états de Lutte contre la Sécheresse dans le Sahel, 2016 Landscapes of West Africa—A window on a changing world.

Codron, J., D. Codron, M. Sponheimer, K. Kirkman, K. J. Duffy *et al.*, 2012 Stable isotope series from elephant ivory reveal lifetime histories of a true dietary generalist. Proc. R. Soc. B Biol. Sci. 279: 2433–2441.

Codron, J., K. Kirkman, K. J. Duffy, M. Sponheimer, J. A. Lee-Thorp *et al.*, 2013 Stable isotope turnover and variability in tail hairs of captive and free-ranging African elephants (*Loxodonta africana*) reveal dietary niche differences within populations. Can. J. Zool. 91: 124–134.

Collier, S., and J. P. White, 1976 Get them young? Age and sex inferences on animal domestication in archaeology. Am. Antiq. 41: 96–102.

Comstock, K. E., N. Georgiadis, J. Pecon-Slattery, A. L. Roca, E. A. Ostrander *et al.*, 2002 Patterns of molecular genetic variation among African elephant populations. Mol. Ecol. 11: 2489–2498.

Corander, J., P. Waldmann, and M. J. Sillanpää, 2003 Bayesian analysis of genetic differentiation between populations. Genetics 163: 367–374.

Cornuet, J. M., and G. Luikart, 1996 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144: 2001–2014.

Cornwell, W. K., I. Wright, J. Turner, V. Maire, M. Barbour *et al.*, 2016 A global dataset of leaf delta 13C values. Sci. Data.

Coutu, A. N., 2019 Historic molecules connect the past to modern conservation, pp. 208–225 in *The Oxford Handbook of Historical Ecology and Applied Archaeology*, edited by Daryl Stump and Christian Isendahl. Newcastle University.

Coutu, A. N., J. Lee-Thorp, M. J. Collins, and P. J. Lane, 2016 Mapping the elephants of the 19th century East African ivory trade with a multi-isotope approach. PLoS One 11: e0163606.

Cui, Z., M. Hui, Y. Liu, C. Song, X. Li *et al.*, 2015 High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab *Eriocheir sinensis*. Heredity (Edinb). 115: 206.

Cui, Y., J. Lindo, C. E. Hughes, J. W. Johnson, A. G. Hernandez *et al.*, 2013 Ancient DNA analysis of mid-holocene individuals from the Northwest Coast of North America reveals different evolutionary paths for mitogenomes. PLoS One 8: e66948.

Cushman, S. A., K. S. McKelvey, J. Hayden, and M. K. Schwartz, 2006 Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. 168: 486–499.

d'Errico, F., and M. Vanhaeren, 2002 Criteria for identifying red deer age and sex from their canines. Application to Upper Palaeolithic and Mesolithic ornaments.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada, 2012 jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9: 772.

Debruyne, R., 2005 A case study of apparent conflict between molecular phylogenies: the interrelationships of African elephants. Cladistics 21: 31–50.

Debruyne, R., A. Van Holt, V. Barriel, and P. Tassy, 2003 Status of the so-called African pygmy elephant (*Loxodonta pumilio* (Noack 1906)): phylogeny of cytochrome b and mitochondrial control region sequences. C. R. Biol. 326: 687–697.

Dobson, J. E., E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley, 2000 LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote Sensing 66: 849–857.

Dublin, H. T., 2003 *IUCN/SSC AfESG guidelines for the in situ translocation of the African elephant for conservation purposes*. IUCN.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Edwards, L. J., K. E. Muller, R. D. Wolfinger, B. F. Qaqish, and O. Schabenberger, 2008 An R2 statistic for fixed effects in the linear mixed model. Stat. Med. 27: 6137–6157.

Eggert, L. S., C. A. Rasner, and D. S. Woodruff, 2002 The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. Proc. R. Soc. London. Ser. B Biol. Sci. 269: 1993–2006.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee *et al.*, 2011 A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17: 43–57.

Enk, J., A. Devault, R. Debruyne, C. E. King, T. Treangen *et al.*, 2011 Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome Biol. 12: R51.

Excoffier, L., and H. E. L. Lischer, 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564–567.

Finch, T. M., N. Zhao, D. Korkin, K. H. Frederick, and L. S. Eggert, 2014 Evidence of positive selection in mitochondrial complexes I and V of the African elephant. PLoS One 9: e92587.

Fishlock, V., and P. C. Lee, 2013 Forest elephants: fission–fusion and social arenas. Anim. Behav. 85: 357–363.

Fitzpatrick, J. L., and J. P. Evans, 2009 Reduced heterozygosity impairs sperm quality in endangered mammals. Biol. Lett. rsbl. 2008.0734.

de Flamingh, A., A. L. Roca, and R. J. van Aarde, 2018 Origin and phylogeography of African savannah elephants (*Loxodonta africana*) in Kruger and nearby parks in southern Africa. Conserv. Genet. 19: 155-167.

de Flamingh, A., A. L. Roca, and R. J. van Aarde, 2017 Origin and phylogeography of African savannah elephants *(Loxodonta africana)* in Kruger and nearby parks in southern Africa. Conserv. Genet. 1–13.

de Flamingh, A., C. L. Sole, and R. J. van Aarde, 2015 Genetic evidence for spatial structuring in a continuous African elephant (*Loxodonta africana*) population. Conserv. Genet. 16: 613-623.

de Flamingh, A., C. L. Sole, and R. J. van Aarde, 2014 Microsatellite repeat motif and amplicon length affect amplification success of degraded faecal DNA. Conserv. Genet. Resour. 6: 503-505.

Frade, F., 1955 Orde des Proboscidiens (Proboscidea Illiger, 1811). In Triaté de Zoologie: Anatomie, Systématique, Biologie (ed P.-P. Grassé). Paris: Masson & Cie.

Fridolfsson, A.-K., and H. Ellegren, 1999 A simple and universal method for molecular sexing of non-ratite birds. J. avian Biol. 116–121.

Fritze, R. H., 2002 *New worlds: the great voyages of discovery, 1400-1600*. Praeger Publishers.

Garant, D., S. E. Forde, and A. P. Hendry, 2007 The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct. Ecol. 21: 434–443.

Garfield, R., 1992 *A history of São Tomé Island, 1470-1655: the key to Guinea*. Edwin Mellen Pr.

Gibbon, V., M. Paximadis, G. Štrkalj, P. Ruff, and C. Penny, 2009 Novel methods of molecular sex identification from skeletal tissue using the amelogenin gene. Forensic Sci. Int. Genet. 3: 74–79.

Gilbert, M. T. P., D. I. Drautz, A. M. Lesk, S. Y. W. Ho, J. Qi *et al.*, 2008 Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc. Natl. Acad. Sci. 105: 8327–8332.

Graham, M. D., I. Douglas-Hamilton, W. M. Adams, and P. C. Lee, 2009 The movement of African elephants in a human-dominated land-use mosaic. Anim. Conserv. 12: 445–455.

Grainger, M., R. Van Aarde, and I. Whyte, 2005 Landscape heterogeneity and the use of space by elephants in the Kruger National Park, South Africa. Afr. J. Ecol. 43: 369–375.

Griffiths, R., M. C. Double, K. Orr, and R. J. G. Dawson, 1998 A DNA test to sex most birds. Mol. Ecol. 7: 1071–1075.

Groves, C. P., 2000 What are the elephants of West Africa? Elephant 2: 5.

Groves, C. P., and P. Grubb, 2000 Do *Loxodonta cyclotis* and *L. africana* interbreed? Elephant 2: 4.

Grubb, P., C. P. Groves, J. P. Dudley, and J. Shoshani, 2000 Living African elephants belong to two species: *Loxodonta africana* (Blumenbach, 1797) and *Loxodonta cyclotis* (Matschie, 1900). Elephant 2: 3.

Guillot, G., F. Mortier, and A. Estoup, 2005 GENELAND: a computer package for landscape genetics. Mol. Ecol. Resour. 5: 712–715.

Gurka, M. J., 2006 Selecting the best linear mixed model under REML. Am. Stat. 60: 19–26.

Hall-Martin, A. J., 1992 Distribution and status of the African elephant Loxodonta africana in South Africa, 1652-1992. Koedoe 35: 65–88.

Hamilton, W. J., R. L. Tilson, and L. G. Frank, 1986 Sexual monomorphism in spotted hyenas, *Crocuta crocuta*. Ethology 71: 63–73.

Hanks, J., 2003 Transfrontier Conservation Areas (TFCAs) in Southern Africa: their role in conserving biodiversity, socioeconomic development and promoting a culture of peace. J. Sustain. For. 17: 127–148.

Hauf, J., P. J. Waddell, N. Chalwatzis, U. Joger, and F. K. Zimmermann, 2000 The complete mitochondrial genome sequence of the African elephant (*Loxodonta africana*), phylogenetic relationships of Proboscidae to other mammals, and D-loop heteroplasmy. ZOOLOGY-JENA- 102: 184–195.

Hauptmann, A., G. Schneider, and C. Bartels, 2016 The shipwreck of Bom Jesus, AD 1533: fugger copper in Namibia. J. African Archaeol. 14: 184–207.

Heaton, T. H. E., J. C. Vogel, G. von La Chevallerie, and G. Collett, 1986 Climatic influence on the isotopic composition of bone nitrogen. Nature 322: 822.

Hedrick, P. W., and A. Garcia-Dorado, 2016 Understanding inbreeding depression, purging, and genetic rescue. Trends Ecol. Evol. 31: 940–952.

Hellevang, H., and P. Aagaard, 2015 Constraints on natural global atmospheric CO 2 fluxes from 1860 to 2010 using a simplified explicit forward model. Sci. Rep. 5: 17352.

Heyer, E., R. Chaix, S. Pavard, and F. Austerlitz, 2012 Sex-specific demographic behaviours that shape human genomic variation. Mol. Ecol. 21: 597–612.

Hoare, R. E., 1999 Determinants of human–elephant conflict in a land-use mosaic. J. Appl. Ecol. 36: 689–700.

Hoare, R. E., and J. T. Du Toit, 1999 Coexistence between people and elephants in African savannas. Conserv. Biol. 13: 633–639.

Hoban, S., 2018 Integrative conservation genetics: prioritizing populations using climate predictions, adaptive potential and habitat connectivity. Mol. Ecol. Resour. 18: 14–17.

Hofreiter, M., V. Jaenicke, D. Serre, A. von Haeseler, and S. Pääbo, 2001 DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Res. 29: 4793–4799.

Höss, M., P. Jaruga, T. H. Zastawny, M. Dizdaroglu, and S. Paabo, 1996 DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res. 24: 1304–1307.

Huber, S., U. Bruns, and W. Arnold, 2002 Sex determination of red deer using polymerase chain reaction of DNA from feces. Wildl. Soc. Bull. 208–212.

Hunter, R. D., R. N. Fisher, and K. R. Crooks, 2003 Landscape-level connectivity in coastal southern California, USA, as assessed through carnivore habitat suitability. Nat. Areas J. 23: 302–314.

Ijoma, J. O., 1982 Portuguese activities in West Africa before 1600: the consequences. Transafrican J. Hist. 11: 136–146.

Ishibashi, H., T. Takeuchi, I. Whyte, and H. Koike, 1999 δ15N and δ13C measurements from the African elephant, *Loxodonta africana*, used for ivory sourcing.

Ishida, Y., Y. Demeke, P. J. van Coeverden de Groot, N. J. Georgiadis, K. E. A. Leggett *et al.*, 2011 Distinguishing forest and savanna African elephants using short nuclear DNA sequences. J. Hered. 102: 610–616.

Ishida, Y., N. J. Georgiadis, T. Hondo, and A. L. Roca, 2013 Triangulating the provenance of African elephants using mitochondrial DNA. Evol. Appl. 6: 253–265.

Jackson, T. P., S. Mosojane, S. M. Ferreira, and R. J. van Aarde, 2008 Solutions for elephant *Loxodonta africana* crop raiding in northern Botswana: moving away from symptomatic approaches. Oryx 42: 83–91.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara, 2008 Hole-filled SRTM for the globe Version 4. available from CGIAR-CSI SRTM 90m Database.

Jobling, M. A., and P. Gill, 2004 Encoded evidence: DNA in forensic analysis. Nat. Rev. Genet. 5: 739.

Johnson, M. B., S. L. Clifford, B. Goossens, S. Nyakaana, B. Curran *et al.*, 2007 Complex phylogeographic history of central African forest elephants and its implications for taxonomy. BMC Evol. Biol. 7: 244.

Jombart, T., 2008 adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405.

Jónsson, H., A. Ginolhac, M. Schubert, P. L. F. Johnson, and L. Orlando, 2013 mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29: 1682–1684.

Kabir, M., H. A. Shoaib Hameed, L. Bosso, J. U. Din, R. Bischof *et al.*, 2017 Habitat suitability and movement corridors of grey wolf (*Canis lupus*) in Northern Pakistan. PLoS One 12.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald, 2014 Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: e281.

Keeley, A. T. H., P. Beier, and J. W. Gagnon, 2016 Estimating landscape resistance from habitat suitability: effects of data source and nonlinearities. Landsc. Ecol. 31: 2151–2162.

Keeley, A. T. H., P. Beier, B. W. Keeley, and M. E. Fagan, 2017 Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. Landsc. Urban Plan. 161: 90–102.

Klinga, P., M. Mikoláš, P. Smolko, M. Tejkal, J. Höglund *et al.*, 2019 Considering landscape connectivity and gene flow in the Anthropocene using complementary landscape genetics and habitat modelling approaches. Landsc. Ecol. 34: 521–536.

Van Klinken, G. J., 1999 Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. 26: 687–695.

Krause, J., P. H. Dear, J. L. Pollack, M. Slatkin, H. Spriggs *et al.*, 2006 Multiplex amplification of the mammoth mitochondrial genome and the evolution of *Elephantidae*. Nature 439: 724–727.

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357.

Larkin, J. L., D. S. Maehr, T. S. Hoctor, M. A. Orlando, and K. Whitney, 2004 Landscape linkages and conservation planning for the black bear in west-central Florida, pp. 23–34 in *Animal Conservation forum*, Cambridge University Press.

Lawrence, A. W., 1963 *Trade castles & forts of West Africa*. London: J. Cape.

Laws, R. M., 1970 Elephants as agents of habitat and landscape change in East Africa. Oikos 1–15.

Leigh, J. W., and D. Bryant, 2015 Popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6: 1110–1116.

Leos-Barajas, V., and T. Michelot, 2018 An Introduction to Animal Movement Modeling with Hidden Markov Models using Stan for Bayesian Inference. arXiv Prepr. arXiv1806.10639.

Leroy, B., C. N. Meynard, C. Bellard, and F. Courchamp, 2016 virtualspecies, an R package to generate virtual species distributions. Ecography (Cop.). 39: 599–607.

Li, G., B. W. Davis, E. Eizirik, and W. J. Murphy, 2016 Phylogenomic evidence for ancient hybridization in the genomes of living cats (*Felidae*). Genome Res. 26: 1–11.

Li, H., and R. Durbin, 2010 Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26: 589–595.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Lindenmayer, D., and J. Fischer, 2007 Landscape modification and habitat fragmentation: a synthesis. Glob. Ecol. Biogeogr. 16: 265–280.

Lindgreen, S., 2012 AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res. Notes 5: 337.

Loarie, S. R., R. J. van Aarde, and S. L. Pimm, 2009 Fences and artificial water affect African savannah elephant movement patterns. Biol. Conserv. 142: 3086–3098.

Lonsdorf, E. V, K. E. Anderson, M. A. Stanton, M. Shender, M. R. Heintz *et al.*, 2014 Boys will be boys: sex differences in wild infant chimpanzee social interactions. Anim. Behav. 88: 79–83.

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin, 1998 Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89: 238–247.

Maikaew, U., W. Wajjwalku, V. Siripholwat, S. Thongtipsiridech, S. Mahasawankul *et al.*, 2011 The Study of Nucleotides Arrangement of Mitochondrial DNA in Thai Elephants (*Elephas maximus indicus*). KKU Vet. J. 17: 11–21.

Maisels, F., S. Strindberg, S. Blake, G. Wittemyer, J. Hart *et al.*, 2013 Devastating decline of forest elephants in Central Africa. PLoS One 8: e59469.

Mapaure, I. N., and B. M. Campbell, 2002 Changes in miombo woodland cover in and around Sengwa Wildlife Research Area, Zimbabwe, in relation to elephants and fire. Afr. J. Ecol. 40: 212–219.

Mark, P., and J. da Silva Horta, 2013 *The forgotten diaspora: Jewish communities in West Africa and the making of the Atlantic world*. Cambridge University Press.

Mashintonio, A. F., S. L. Pimm, G. M. Harris, R. J. Van Aarde, and G. J. Russell, 2014 Data-driven discovery of the spatial scales of habitat choice by elephants. PeerJ 2: e504.

Mateo-Sánchez, M. C., N. Balkenhol, S. Cushman, T. Pérez, A. Domínguez *et al.*, 2015 Estimating effective landscape distances and movement corridors: comparison of habitat and genetic data. Ecosphere 6: 1–16.

McDonald, J. H., 2009 *Handbook of biological statistics*. Sparky House Publishing Baltimore, MD.

McKelvey, K. S., J. von Kienast, K. B. Aubry, G. M. Koehler, B. T. Maletzke *et al.*, 2006 DNA analysis of hair and scat collected along snow tracks to document the presence of Canada lynx. Wildl. Soc. Bull. 34: 451–455.

McRae, B. H., 2006 Isolation by resistance. Evolution (N. Y). 60: 1551–1561.

McRae, B. H., and P. Beier, 2007 Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci.

Merow, C., M. J. Smith, and J. A. Silander Jr, 2013 A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography (Cop.). 36: 1058–1069.

Van der Merwe, N. J., J. A. Lee-Thorp, J. F. Thackeray, A. Hall-Martin, F. J. Kruger *et al.*, 1990 Source-area determination of elephant ivory by isotopic analysis. Nature 346: 744.

Van der Merwe, N. J., and E. Medina, 1991 The canopy effect, carbon isotope ratios and foodwebs in Amazonia. J. Archaeol. Sci. 18: 249–259.

Meyer, M., E. Palkopoulou, S. Baleka, M. Stiller, K. E. H. Penkman *et al.*, 2017 Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. Elife 6: e25413.

Michelmore, F., K. Beardsley, R. F. W. Barnes, and I. Douglas-Hamilton, 1994 A model illustrating the changes in forest elephant numbers caused by poaching. Afr. J. Ecol. 32: 89–99.

Milanesi, P., R. Holderegger, R. Caniglia, E. Fabbri, M. Galaverni *et al.*, 2017 Expert-based versus habitat-suitability models to develop resistance surfaces in landscape genetics. Oecologia 183: 67–79.

Miller, M. P., 2005 Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. J. Hered. 96: 722–724.

Miller, E. J., M. D. B. Eldridge, K. D. Morris, K. R. Zenger, and C. A. Herbert, 2011 Genetic consequences of isolation: island tammar wallaby (*Macropus eugenii*) populations and the conservation of threatened species. Conserv. Genet. 12: 1619–1631.

Mittnik, A., C. C. Wang, J. Svoboda, and J. Krause, 2016 A molecular approach to the sexing of the triple burial at the upper paleolithic site of Dolní Vêstonice. PLoS One 11: 1–9.

Mondol, S., I. Moltke, J. Hart, M. Keigwin, L. Brown *et al.*, 2015 New evidence for hybrid zones of forest and savanna elephants in Central and West Africa. Mol. Ecol. 24: 6134–6147.

Monti, F., F. Delfour, V. Arnal, S. Zenboudji, O. Duriez *et al.*, 2018 Genetic connectivity among osprey populations and consequences for conservation: philopatry versus dispersal as key factors. Conserv. Genet. 19: 839–851.

Mowa, E., 2018 Oranjemund shipwreck ivory: historical analysis on the prospective geographic origin. Eur. Cent. Res. Train. Dev. UK 6: 48–69.

Murata, Y., T. Yonezawa, I. Kihara, T. Kashiwamura, Y. Sugihara *et al.*, 2009 Chronology of the extant African elephant species and case study of the species identification of the small African elephant with the molecular phylogenetic method. Gene 441: 176–186.

Murphy, B. P., and D. M. J. S. Bowman, 2006 Kangaroo metabolism does not cause the relationship between bone collagen δ15N and water availability. Funct. Ecol. 20: 1062–1069.

Murphy, B. P., and D. M. J. S. Bowman, 2009 The carbon and nitrogen isotope composition of Australian grasses in relation to climate. Funct. Ecol. 23: 1040–1049.

Murphy, S. M., J. S. Laufenberg, J. D. Clark, M. Davidson, J. L. Belant *et al.*, 2018 Genetic diversity, effective population size, and structure among black bear populations in the Lower Mississippi Alluvial Valley, USA. Conserv. Genet. 19: 1055–1067.

Nakamura, M., 2009 Sex determination in amphibians, pp. 271–282 in *Seminars in cell & developmental biology*, Elsevier.

Nei, M., F. Tajima, and Y. Tateno, 1983 Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19: 153–170.

Newitt, M., 2010 *The Portuguese in West Africa, 1415–1670: a documentary history*. Cambridge University Press.

Newmark, W. D., 2008 Isolation of African protected areas. Front. Ecol. Environ. 6: 321–328.

Nyakaana, S., and P. Arctander, 1999 Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. Mol. Ecol. 8: 1105–1115.

Nyakaana, S., P. Arctander, and H. R. Siegismund, 2002 Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. Heredity (Edinb). 89: 90.

Okello, J. B. A., C. Masembe, H. B. Rasmussen, G. Wittemyer, P. Omondi *et al.*, 2008 Population genetic structure of savannah elephants in Kenya: conservation and management implications. J. Hered. 99: 443–452.

Orelien, J. G., and L. J. Edwards, 2008 Fixed-effect variable selection in linear mixed models using R2 statistics. Comput. Stat. Data Anal. 52: 1896–1907.

Orton, R. W., D. B. Tucker, J. S. Harrison, and L. D. McBrayer, 2020 Spatial and temporal patterns of genetic diversity in a fragmented and transient landscape. Evol. Ecol. 1–17.

Osborn, F. V, and G. E. Parker, 2003 Linking two elephant refuges with a corridor in the communal lands of Zimbabwe. Afr. J. Ecol. 41: 68–74.

Palkopoulou, E., M. Lipson, S. Mallick, S. Nielsen, N. Rohland *et al.*, 2018 A comprehensive genomic history of extinct and living elephants. Proc. Natl. Acad. Sci. 115: E2566–E2574.

Peakall, R. O. D., and P. E. Smouse, 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Resour. 6: 288–295.

Pedersen, J. M., S. E. Glickman, L. G. Frank, and F. A. Beach, 1990 Sex differences in the play behavior of immature spotted hyenas, *Crocuta crocuta*. Horm. Behav. 24: 403–420.

Peery, M. Z., R. Kirby, B. N. Reid, R. Stoelting, E. Doucet-Bëer *et al.*, 2012 Reliability of genetic bottleneck tests for detecting recent population declines. Mol. Ecol. 21: 3403–3418.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward, 2016 High-resolution mapping of global surface water and its long-term changes. Nature 540: 418–422.

Pelletier, F., G. Turgeon, A. Bourret, D. Garant, and M.-H. St-Laurent, 2019 Genetic structure and effective size of an endangered population of woodland caribou. Conserv. Genet. 20: 203–213.

Petit, R. J., and L. Excoffier, 2009 Gene flow and species delimitation. Trends Ecol. Evol. 24: 386–393.

Pettorelli, N., J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker *et al.*, 2005 Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol. Evol. 20: 503–510.

Phillips, S. J., and M. Dudík, 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography (Cop.). 31: 161–175.

Phillips, S. J., M. Dudík, and R. E. Schapire, 2017 Maxent software for modeling species niches and distributions (Version 3.4. 1). Biodivers. Informatics.

"Picard Toolkit." 2019. Broad Institute, G. R. http://broadinstitute. github. io/picard/. B. I. No Title.

Pienaar, U. D. V, 1963 The large mammals of the Kruger National Park-their distribution and present-day status. Koedoe 6: 1–37.

Pinter-Wollman, N., 2012 Human–elephant conflict in Africa: the legal and political viability of translocations, wildlife corridors, and transfrontier parks for large mammal conservation. J. Int. Wildl. Law Policy 15: 152–166.

Poulsen, J. R., C. Rosin, A. Meier, E. Mills, C. L. Nuñez *et al.*, 2018 Ecological consequences of forest elephant declines for Afrotropical forests. Conserv. Biol. 32: 559–567.

Pullinger, M. G., and C. J. Johnson, 2010 Maintaining or restoring connectivity of modified landscapes: evaluating the least-cost path model with multiple sources of ecological information. Landsc. Ecol. 25: 1547–1560.

Quincey, D., G. Carle, V. Alunni, and G. Quatrehomme, 2013 Difficulties of sex determination from forensic bone degraded DNA: A comparison of three methods. Sci. Justice 53: 253–260.

Quinn, A. E., R. S. Radder, S. D. Sarre, A. Georges, T. Ezaz *et al.*, 2009 Isolation and development of a molecular sex marker for *Bassiana duperreyi*, a lizard with XX/XY sex chromosomes and temperature-induced sex reversal. Mol. Genet. genomics 281: 665–672.

R-Development-Core-Team, 2017 R: A language and environment for statistical computing.

R Core Team, 2019 R Core Team (2019). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. URL http//www. R-project. org/., page R Found. Stat. Comput.

Rambaut, A., and A. Drummond, 2009 *FigTree v1. 3.1*. Institute of Evolutionary Biology, University of Edinburgh .

Rambaut, A., A. J. Drummond, and M. Suchard, 2014 Tracer v1. 6 http://beast. bio. ed. ac. uk. Tracer>(Online 2015, May 29).

Ripple, W. J., T. M. Newsome, C. Wolf, R. Dirzo, K. T. Everatt *et al.*, 2015 Collapse of the world's largest herbivores. Sci. Adv. 1: e1400103.

Robson, A. S., M. J. Trimble, A. Purdon, K. D. Young-Overton, S. L. Pimm *et al.*, 2017 Savanna elephant numbers are only a quarter of their expected values. PLoS One 12: e0175942.

Roca, A. L., N. Georgiadis, J. Pecon-Slattery, and S. J. O'brien, 2001 Genetic evidence for two species of elephant in Africa. Science (80-.). 293: 1473–1477.

Roca, A. L., Y. Ishida, A. L. Brandt, N. R. Benjamin, K. Zhao *et al.*, 2015 Elephant natural history: a genomic perspective. Annu. Rev. Anim. Biosci. 3: 139–167.

Roever, C. L., R. J. van Aarde, and K. Leggett, 2013 Functional connectivity within conservation networks: Delineating corridors for African elephants. Biol. Conserv. 157: 128–135.

Roever, C. L., R. J. Van Aarde, and K. Leggett, 2012 Functional responses in the habitat selection of a generalist mega-herbivore, the African savannah elephant. Ecography (Cop.). 35: 972–982.

Rogaev, E. I., Y. K. Moliaka, B. A. Malyarchuk, F. A. Kondrashov, M. V Derenko *et al.*, 2006 Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. PLoS Biol. 4.

Rogers, T. L., 2005 Determining the sex of human remains through cranial morphology. J. Forensic Sci. 50: 1–8.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling *et al.*, 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61: 539–542.

Roth, H. H., and I. Douglas-Hamilton, 1991 Distribution and status of elephants in West Africa. Mammalia 55: 489–528.

Row, J. R., S. T. Knick, S. J. Oyler-McCance, S. C. Lougheed, and B. C. Fedy, 2017 Developing approaches for linear mixed modeling in landscape genetics through landscape-directed dispersal simulations. Ecol. Evol. 7: 3751–3761.

Royle, J. A., R. B. Chandler, C. Yackulic, and J. D. Nichols, 2012 Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. Methods Ecol. Evol. 3: 545–554.

Sæther, B.-E., E. J. Solberg, and M. Heim, 2003 Effects of altering sex ratio structure on the demography of an isolated moose population. J. Wildl. Manage. 455–466.

Safont, S., A. Malgosa, and M. E. Subirà, 2000 Sex assessment on the basis of long bone circumference. Am. J. Phys. Anthropol. Off. Publ. Am. Assoc. Phys. Anthropol. 113: 317–328.

Schoener, T. W., and A. Schoener, 1980 Densities, sex ratios, and population structure in four species of Bahamian Anolis lizards. J. Anim. Ecol. 19–53.

Schwartz SB, 2007 *Portuguese oceanic expansion, 1400-1800* (C. D. The Economy of the Portuguese Empire. In: Bethencourt F, Ed.). Cambridge: Cambridge University Press; pp. 19–48.

Schwarz, C., R. Debruyne, M. Kuch, E. McNally, H. Schwarcz *et al.*, 2009 New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res. 37: 3215–3229.

Seddon, J. M., and B. Schultz, 2020 Koala Conservation in Queensland, Australia: A Role for Assisted Gene Flow for Genetic Rescue?, pp. 331–349 in *Conservation Genetics in Mammals*, Springer.

Shah, V. B., and B. H. McRae, 2008 Circuitscape: a tool for landscape ecology, pp. 62–66 in *Proceedings of the 7th Python in Science Conference*, SciPy 2008, Pasadena, California, USA.

Shirk, A. J., S. A. Cushman, and E. L. Landguth, 2012 Simulating pattern-process relationships to validate landscape genetic models. Int. J. Ecol. 2012.

Soares, M. de C., 2017 " Por conto e peso": o comércio de marfim no Congo e Loango, séculos XV–XVII. An. do Mus. Paul. História e Cult. Mater. 25: 59–86.

Sowunmi, M. A., 2004 *Aspects of Nigerian coastal vegetation in the Holocene: some recent insights*. Springer.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Stanton, D. W. G., J. Hart, A. Vosper, N. F. Kümpel, J. Wang *et al.*, 2016 Non-invasive genetic identification confirms the presence of the Endangered okapi *Okapia johnstoni* south-west of the Congo River. Oryx 50: 134–137.

Van Strien, M. J., D. Keller, and R. Holderegger, 2012 A new analytical approach to landscape genetic modelling: Least-cost transect analysis and linear mixed models. Mol. Ecol. 21: 4010–4023.

Sukumar, R., 2003 *The living elephants: evolutionary ecology, behaviour, and conservation*. Oxford University Press.

Sullivan, K. M., A. Mannucci, C. P. Kimpton, and P. Gill, 1993 A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of XY homologous gene amelogenin. Biotechniques 15: 636–638.

Taberlet, P., L. P. Waits, and G. Luikart, 1999 Noninvasive genetic sampling: look before you leap. Trends Ecol. Evol. 14: 323–327.

Tang, Q., T. Fung, and F. E. Rheindt, 2019 ResDisMapper: An r package for fine-scale mapping of resistance to dispersal. Mol. Ecol. Resour.

Tchamba, M. N., and P. M. Seme, 1993 Diet and feeding behaviour of the forest elephant in the Santchou Reserve, Cameroon. Afr. J. Ecol. 31: 165–171.

Thornton, J. K., 1998 *Africa and Africans in the making of the Atlantic world, 1400-1800*. Cambridge University Press.

Thornton, J. K., 2007 *Portuguese oceanic expansion, 1400-1800* (T. P. in A. I. F. B. and D. Curto, Ed.). Cambridge: Cambridge University Press; pp. 138–160.

Thouless, C., H. T. Dublin, J. J. Blanc, D. P. Skinner, T. E. Daniel *et al.*, 2016 African elephant status report 2016. Occas. Pap. Ser. IUCN Species Surviv. Comm. 60.

Verbeke, G., and G. Molenberghs, 2000 *Linear mixed models for longitudinal data*. Springer.

Vogel, S. M., B. Lambert, A. C. Songhurst, G. P. McCulloch, A. Lee Stronza *et al.*, 2020 Exploring movement decisions: can Bayesian movement-state models explain crop consumption behaviour in elephants (*Loxodonta africana*)? J. Anim. Ecol.

Vogel, J. C., A. S. Talma, A. J. Hall-Martint, and P. J. Viljoen, 1990 Carbon and nitrogen isotopes in elephants. Am. J. Psychiat 142: 163–170.

Wall, J., I. Douglas-Hamilton, and F. Vollrath, 2006 Elephants avoid costly mountaineering. Curr. Biol. 16: R527–R529.

Wandeler, P., S. Smith, P. A. Morin, R. A. Pettifor, and S. M. Funk, 2003 Patterns of nuclear DNA degeneration over time—a case study in historic teeth samples. Mol. Ecol. 12: 1087–1093.

Wang, G., 2019 Machine learning for inferring animal behavior from location and movement data. Ecol. Inform. 49: 69–76.

Wang, Y.-H., K.-C. Yang, C. L. Bridgman, and L.-K. Lin, 2008 Habitat suitability modelling to correlate gene flow with landscape connectivity. Landsc. Ecol. 23: 989–1000.

Waples, R. S., 2016 Making sense of genetic estimates of effective population size. Mol. Ecol. 25: 4689–4691.

Wasser, S. K., L. Brown, C. Mailand, S. Mondol, W. Clark *et al.*, 2015 Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science 349: 84–87.

Webber, B. L., C. J. Yates, D. C. Le Maitre, J. K. Scott, D. J. Kriticos *et al.*, 2011 Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models. Divers. Distrib. 17: 978–1000.

Werz, B. E. J. S., 2010 Sub-saharan Africa's oldest shipwreck: historical-archeological research of an early modern-era Portuguese merchantman on the Namibian coast. Mar. Mirror 96: 430–442.

White, F., 1983 *The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa by F White*. Unesco, Paris.

Whitehouse, A. M., and E. H. Harley, 2001 Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. Mol. Ecol. 10: 2139–2149.

Whyte, I. J., 2001 Conservation management of the Kruger National Park elephant population: University of Pretoria.

Wittemyer, G., J. M. Northrup, J. Blanc, I. Douglas-Hamilton, P. Omondi *et al.*, 2014 Illegal killing for ivory drives global decline in African elephants. Proc. Natl. Acad. Sci. 111: 13117–13121.

Xavier, M., 1989 *Relações da carreira da Índia*. Publicações Alfa.

Young, K. D., S. M. Ferreira, and R. J. van Aarde, 2009a The influence of increasing population size and vegetation productivity on elephant distribution in the Kruger National Park. Austral Ecol. 34: 329–342.

Young, K. D., S. M. Ferreira, and R. J. Van Aarde, 2009b Elephant spatial use in wet and dry savannas of southern Africa. J. Zool. 278: 189–205.

Ziegler, S., S. Merker, B. Streit, M. Boner, and D. E. Jacob, 2016 Towards understanding isotope variability in elephant ivory to establish isotopic profiling and source-area determination. Biol. Conserv. 197: 154–163.

APPENDIX A: SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 2

Table A.1 Marker names, profiles and characteristics[4] of nine microsatellite loci used in this study.

Marker Name	Locus Name	Repeat Motif	Size (bp)	Ta-°C	No of alleles	HO	HE
LaT08[2]	1	(TAGA)16	166–234	56	13	0.81955	0.86347
Lat13[1]	2	(CATC)21	173–262	56	10	0.62963	0.76649
Lat17[1]	3	(GGAT)15… (GGAT)	324–352	56	8	0.71212	0.80626
Lat24[1]	4	(GGAT)22	203–231	56	8	0.77953	0.84075
FH1[2]	5	(CA)12	81-89	55	6	0.56522	0.61660
FH39[2]	6	(CA)18	232-256	60	12	0.63704	0.78441
FH102[2]	7	(CT)11(CA)14	175-187	60	6	0.51773	0.49862
LA5[3]	8	(CA)13	139–153	52	8	0.56028	0.55996
Lat25[1]	9	(CCAT)15	287–321	52	9	0.66923	0.83710
Mean					8.889	0.65448	0.73041
s.d.					2.421	0.10179	0.13554

[1] Comstock et al., 2000

[2] Archie et al., 2003

[3] Eggert et al., 2000

[4] Number of alleles per locus, Observed Heterozygocity (HO) and Expected Heterozygocity (HE)

Table A.2 The area under the receiver-operator curve (AUC) indicated that gradient-based human presence MaxEnt models performed less well than models that considered raw "ambient population" presence.

Background data	Maxent regularization parameter	AUC for gradient-based human presence	AUC for "ambient population" presence
Entire extent	$\beta = 1$	0.848706	0.8568687
	$\beta = 2$	0.8458197	0.854843
	$\beta = 3$	0.842903	0.8524966
	$\beta = 4$	0.8390468	0.8492671
Use-based background data	$\beta = 1$	0.7529524	0.7860084
	$\beta = 2$	0.7477055	0.7808212
	$\beta = 3$	0.7418662	0.7745906
	$\beta = 4$	0.7354542	0.7644965
Range-based background data	$\beta = 1$	0.7846587	0.7583397
	$\beta = 2$	0.7774159	0.7544933
	$\beta = 3$	0.7703632	0.7505779
	$\beta = 4$	0.7604415	0.7448165

Table A.3 We calculated MLPE models for three resistance surfaces which included linear, slight nonlinear, and pronounced nonlinear transformations of the habitat suitability model (HSM) and four genetic distance (GD) response variables (D_{PS} = 1 minus the proportion of shared alleles; GD_Aldiff = genetic distance as the number of allelic differences between two individuals; GD_euc = genetic distance as the Euclidean distance among a vector of allele frequencies; GD_TotGD = Reynold's genetic distance measure). For each of these resistance surface and genetic distance combinations, we fitted MLPE models that considered as fixed effects geographic Euclidean distance (Geo ED) only, least-cost path distance (LCP) and Geo ED, and resistance distance (CT) and Geo ED, resulting in a comparison of 36 MLPE models in total. Based on AIC, conditional and marginal R^2, we find that CT and Geo ED calculated using the pronounced HSM transformation results in the best predictor of D_{PS} as a proxy of gene flow across the landscape (indicated in **boldface**).

	Linear Transformation			*Slight nonlinear transformation*			*Pronounced nonlinear transformation*		
	AIC	R2m	R2c	AIC	R2m	R2c	AIC	R2m	R2c
D_{PS} ~Geo ED	-30950.4	0.001	0.099	-30950.4	0.001	0.099	-30950.4	0.001	0.099
D_{PS} ~LCP + Geo ED	-30952.4	0.002	0.101	-30952.6	0.002	0.101	-30951.8	0.002	0.101
D_{PS} ~CT + Geo ED	-30952.3	0.002	0.100	-30961.5	0.003	0.102	**-30983.4**	**0.007**	**0.107**
GD_Aldiff ~Geo ED	26920.6	0.001	0.099	26920.6	0.001	0.099	26920.6	0.001	0.099
GD_Aldiff ~LCP + Geo ED	26918.7	0.002	0.101	26918.4	0.002	0.101	26919.2	0.002	0.101
GD_Aldiff ~CT + Geo ED	26918.7	0.002	0.100	26909.5	0.003	0.102	26887.6	0.007	0.107
GD_euc ~Geo ED	113.5	0.001	0.095	113.5	0.001	0.095	113.5	0.001	0.095
GD_euc ~LCP + Geo ED	111.5	0.002	0.097	111.1	0.002	0.097	112	0.002	0.097
GD_euc ~CT + Geo ED	113.6	0.001	0.095	105.8	0.003	0.097	84.4	0.006	0.102
GD_TotGD ~Geo ED	-46927.4	0.001	0.092	-46927.4	0.001	0.092	-46927	0.001	0.092
GD_TotGD ~LCP + Geo ED	-46929.4	0.002	0.094	-46929.8	0.002	0.094	-46929	0.002	0.093
GD_TotGD ~CT + Geo ED	-46926.5	0.001	0.092	-46933.3	0.002	0.094	-46954	0.006	0.098

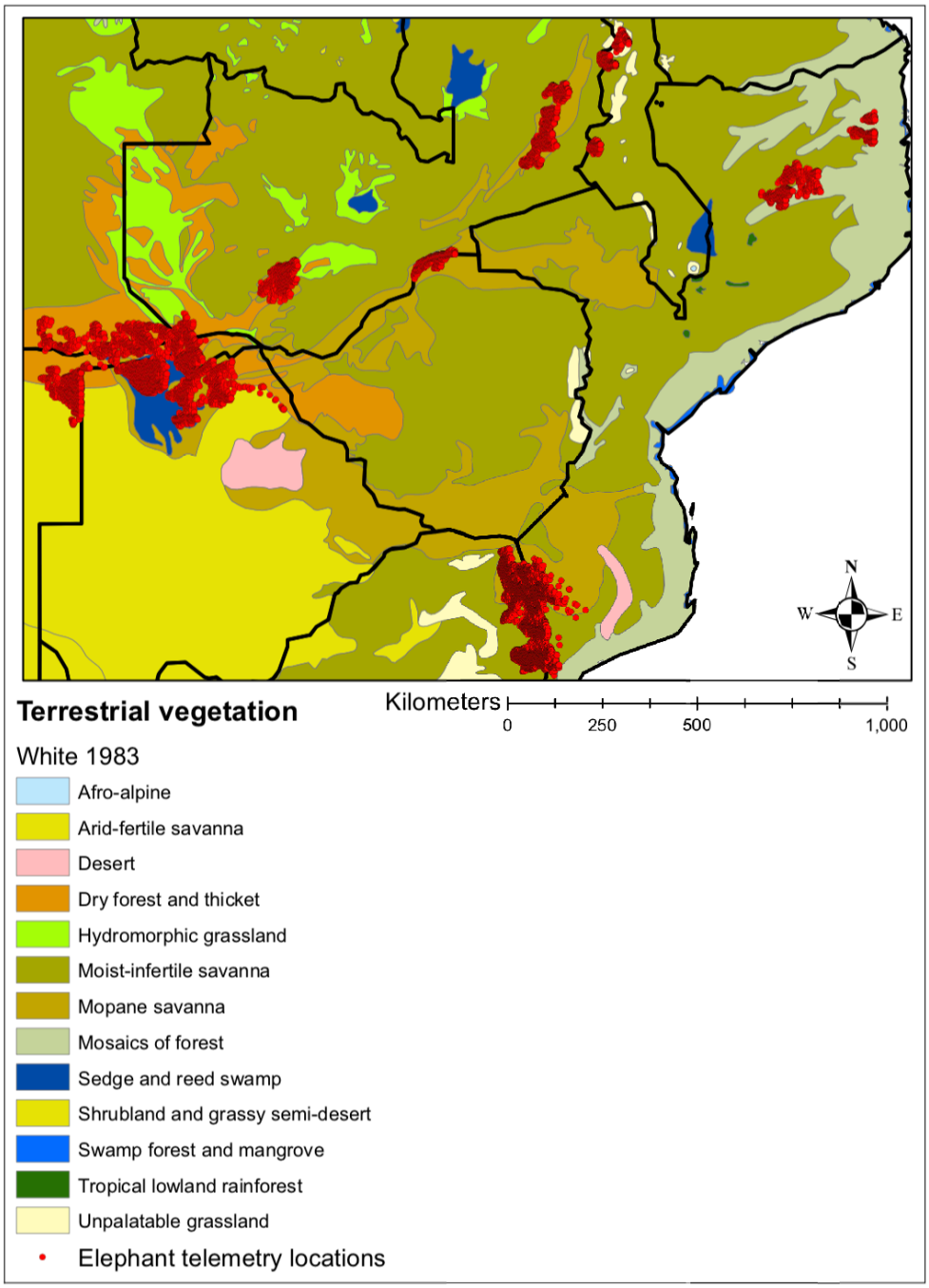

Figure A.1 Our study area includes seven Southern African countries (delineated by black lines) namely Botswana, Malawi, Mozambique, Namibia, South Africa, Zambia and Zimbabwe. Elephant occurrence data (red dots) spans a range of different vegetation classes (White 1983) across the study area.

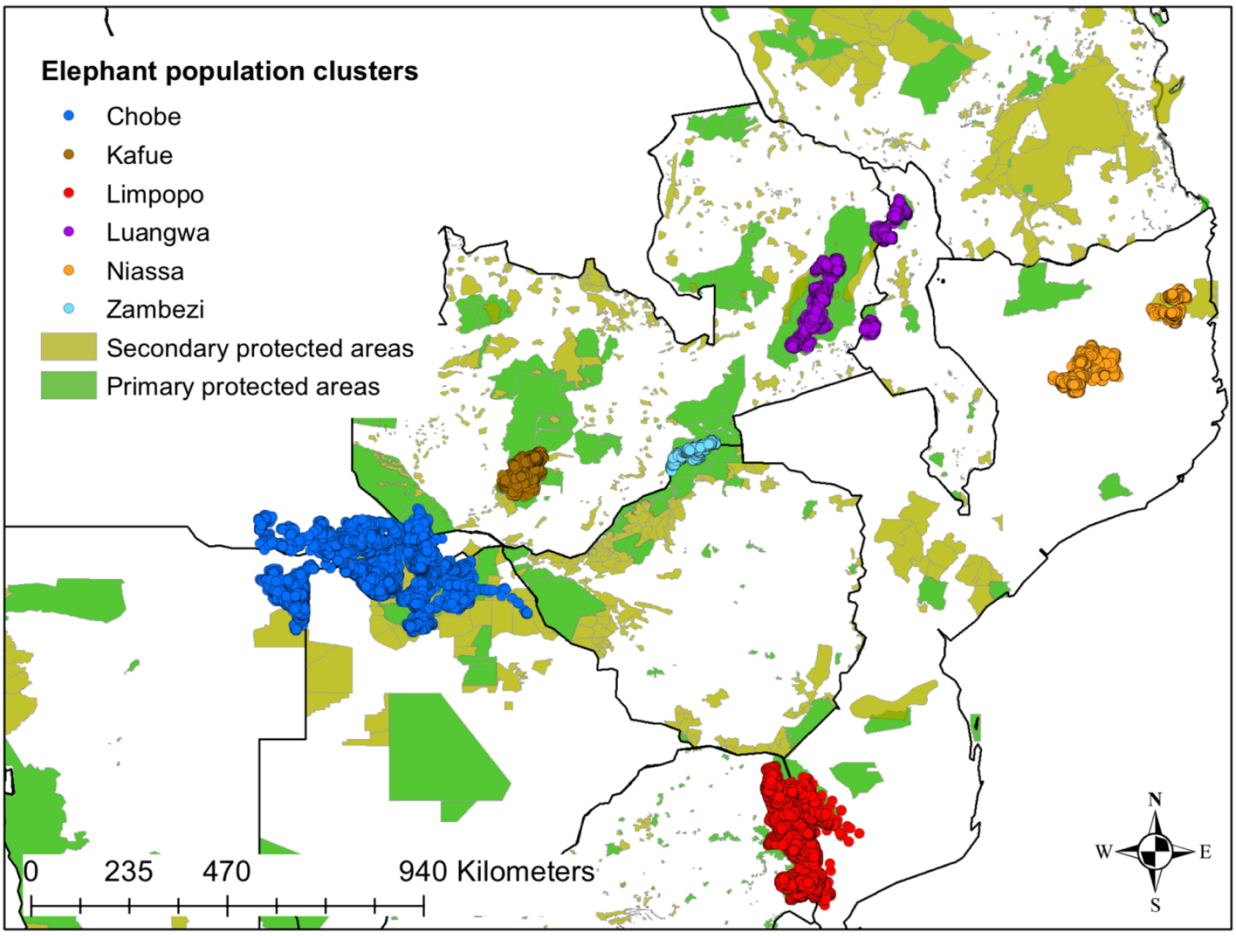

Figure A.2 Spatial location data were from 116 elephants that form part of 6 regional population clusters that include the Chobe, Kafue, Limpopo, Luangwa, Niassa and Zambezi elephant population clusters.

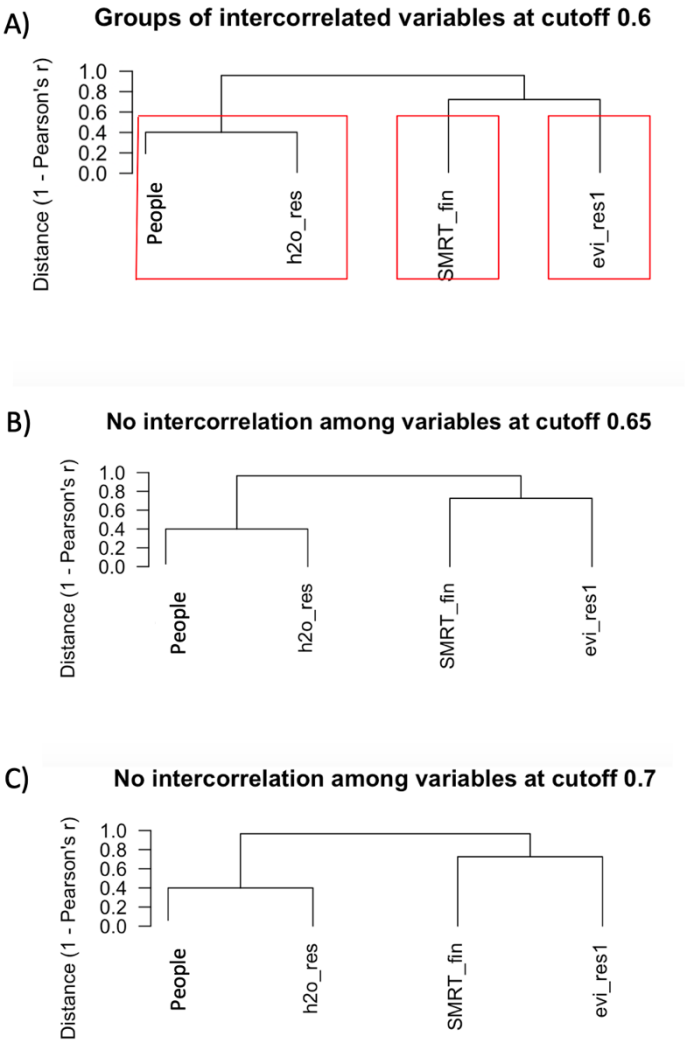

Figure A.3 The transformed gradient-based human distance metric was significantly correlated with our water availability environmental variable at a Pearson's correlation coefficient cut-off of R = 0.6 (panel A – red boxes enclose correlated variables), but no environmental variables were correlated at R = 0.65 (panel B) or R = 0.7 (panel C). Environmental layers included human density as a transformed gradient-based distance (People), water availability as the distance to the closest water source (h2o_res), slope of the landscape (SMRT_fin) and primary productivity (evi_res1).

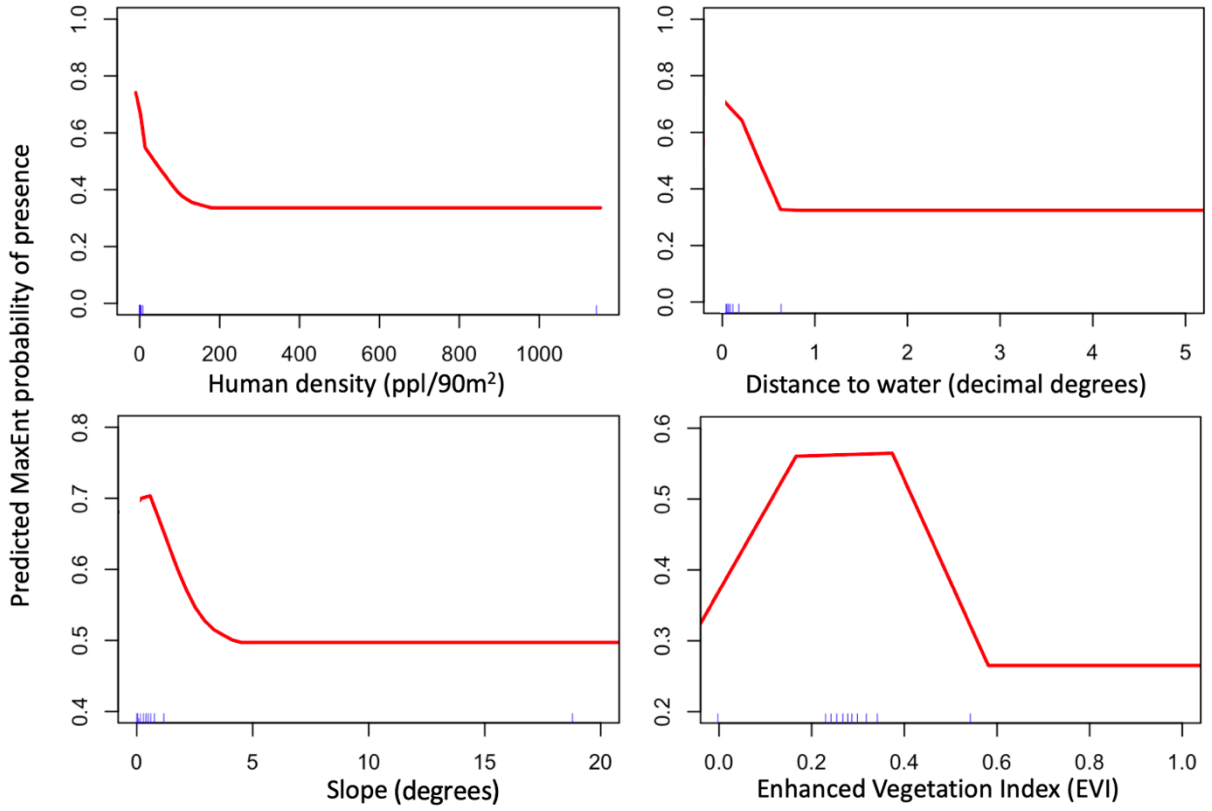

Figure A.4 We found that human density as "ambient population" presence, distance to water and slope mostly contribute to the predictive habitat suitability model (HSM) at low environmental values (low human densities, close to water, and habitats with low slopes), and all three of these environmental variables decreased in their contribution until they reach threshold environmental value beyond which they do not contribute to the predictive model. Intermediate values of primary productivity (EVI) contribute the most to the predictive HSM.

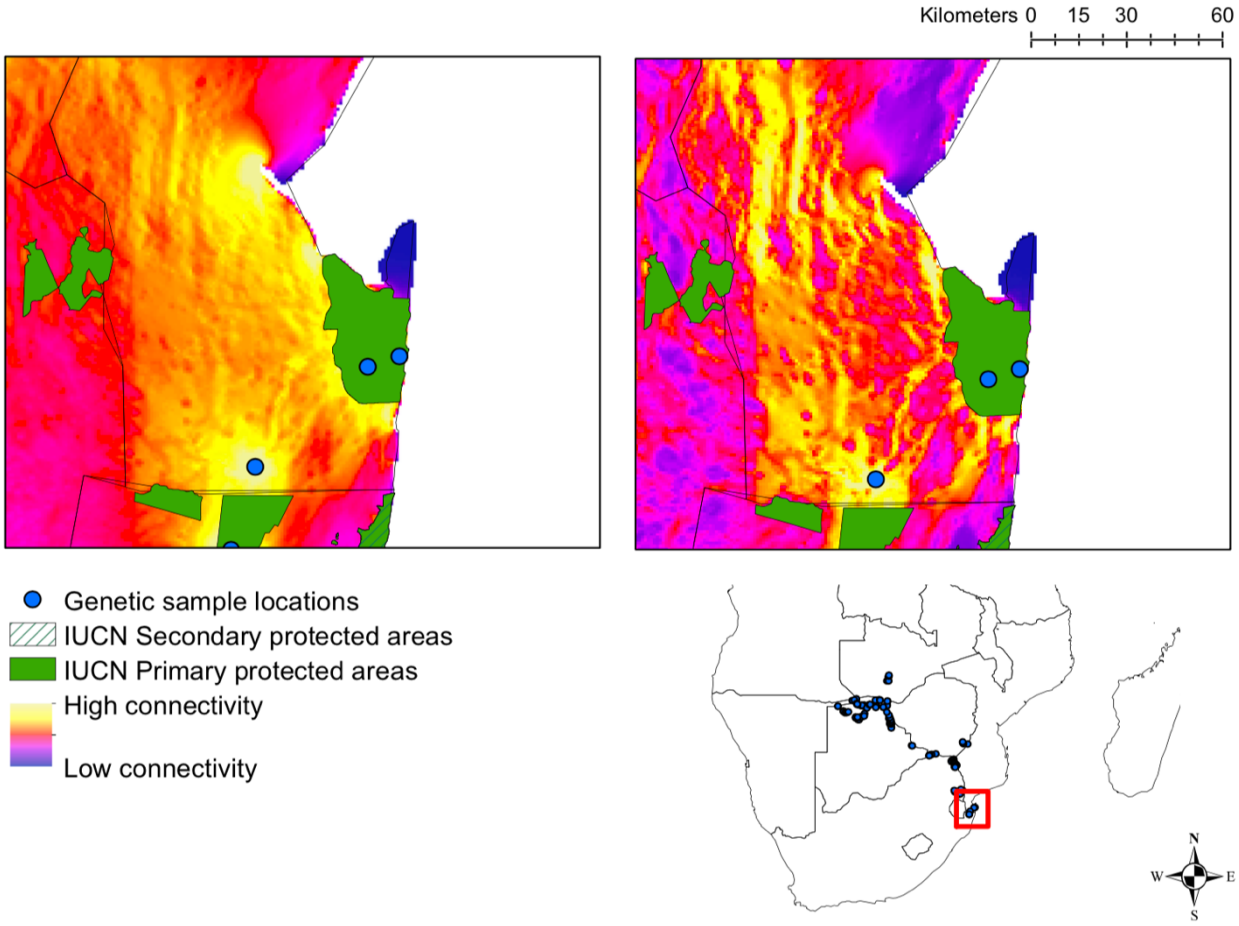

Figure A.5 FLL maps that are based on both spatial and genetic data (right panel) provide more precise delineations of landscape linkages compared to connectivity maps based on only spatial data (left panel), and may therefore be beneficial when delineating fine scale linkages across the landscape, where spatial data alone may not provide a high enough resolution for demarcating linkages as proposed conservation areas.

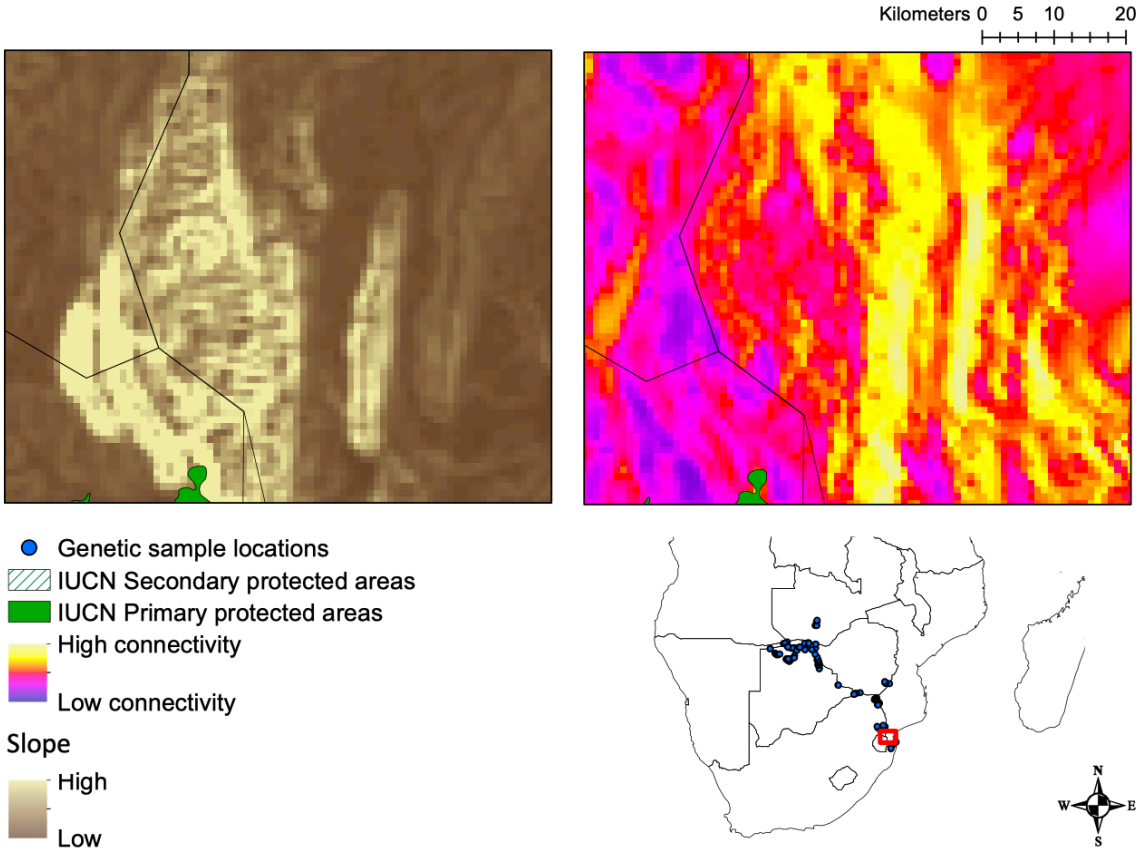

Figure A.6 There are areas in our FLL map where slope seems to be the primary driver of connectivity. For example, areas of high slope (yellow in the left panel) correspond to areas with low landscape connectivity (purple in the right panel).

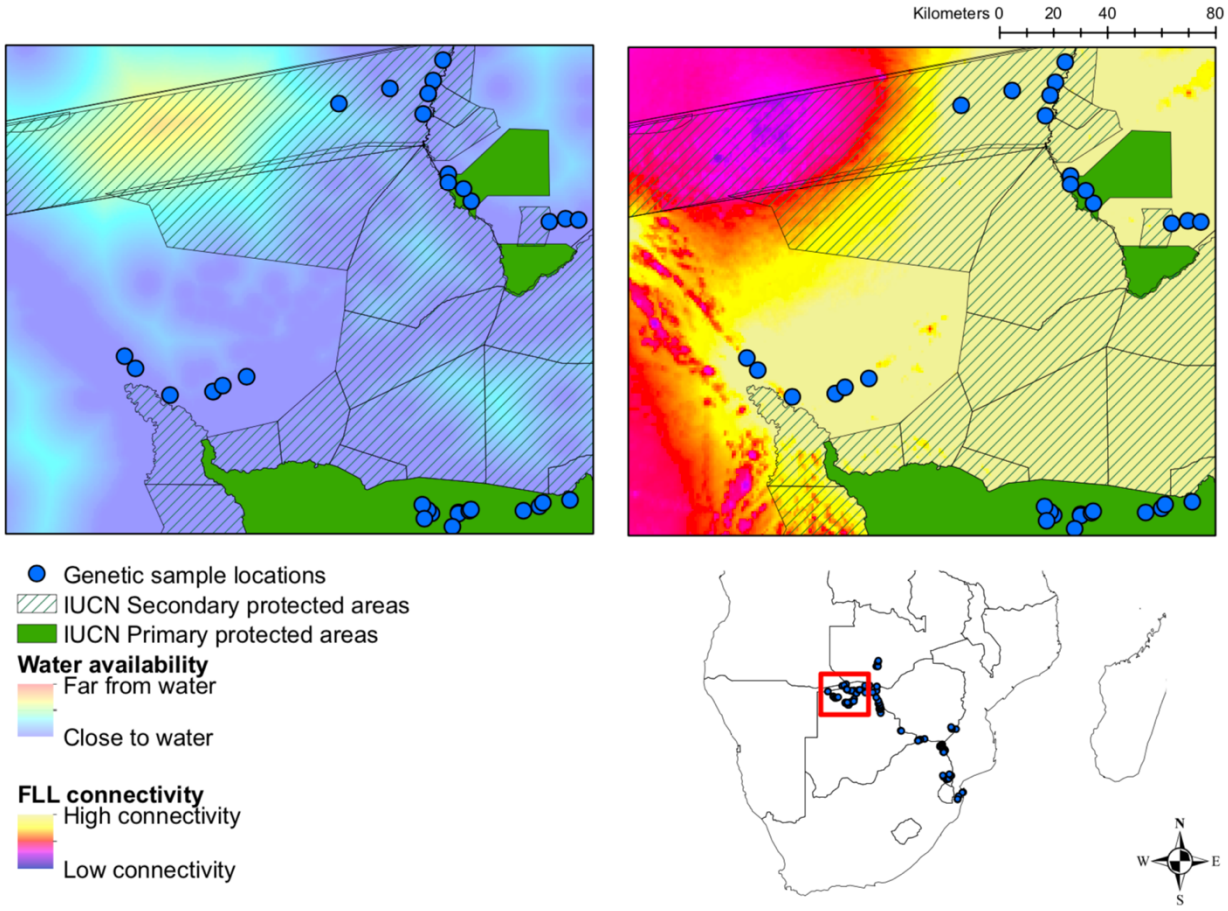

Figure A.7 There are areas in our FLL map where water availability seems to be the primary driver of connectivity. For example, areas that are far from water (yellow and red in the left panel) correspond to areas with low landscape connectivity (purple in the right panel).

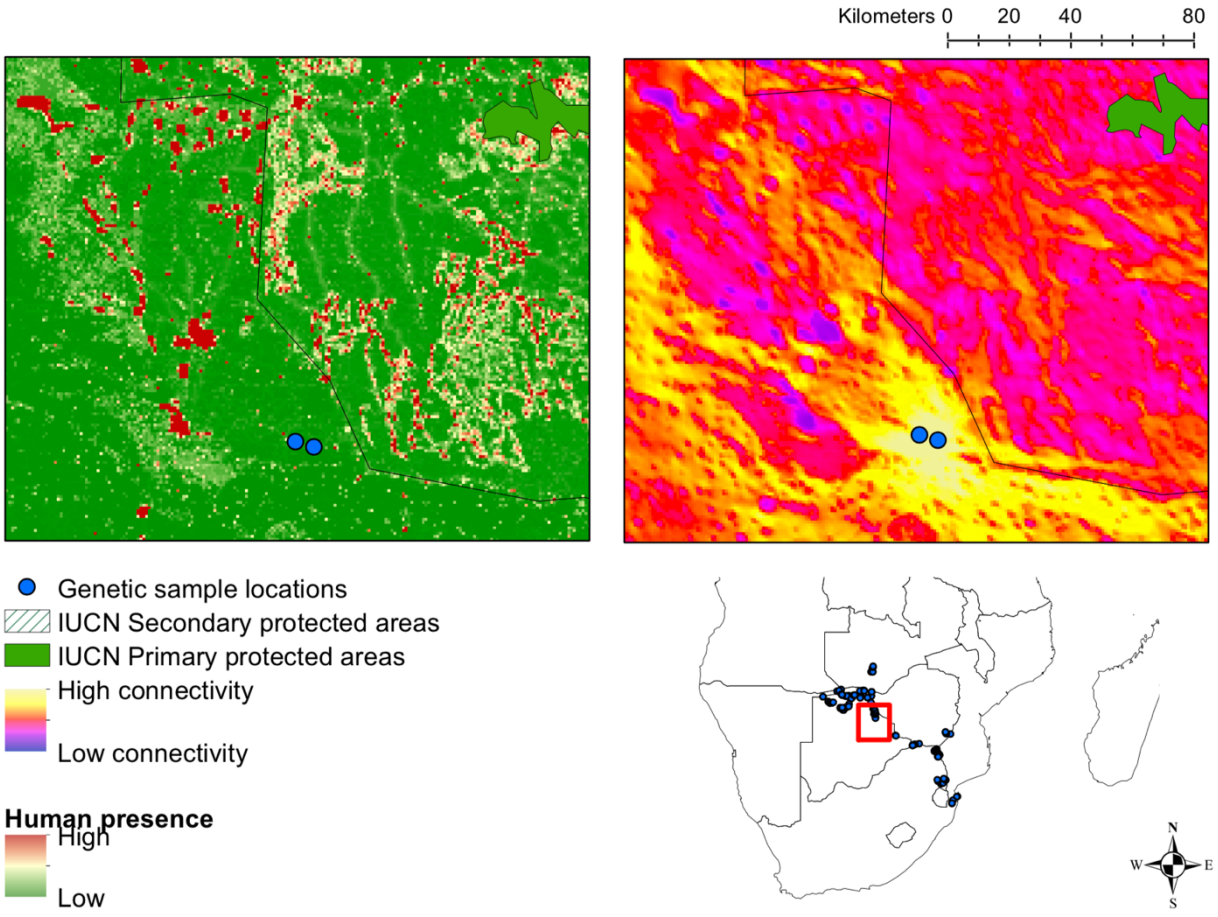

Figure A.8 There are areas in our FLL map where human presence seems to be the primary driver of connectivity. For example, areas of high "ambient population" presence (red in the left panel) correspond to areas with low landscape connectivity (purple in the right panel).

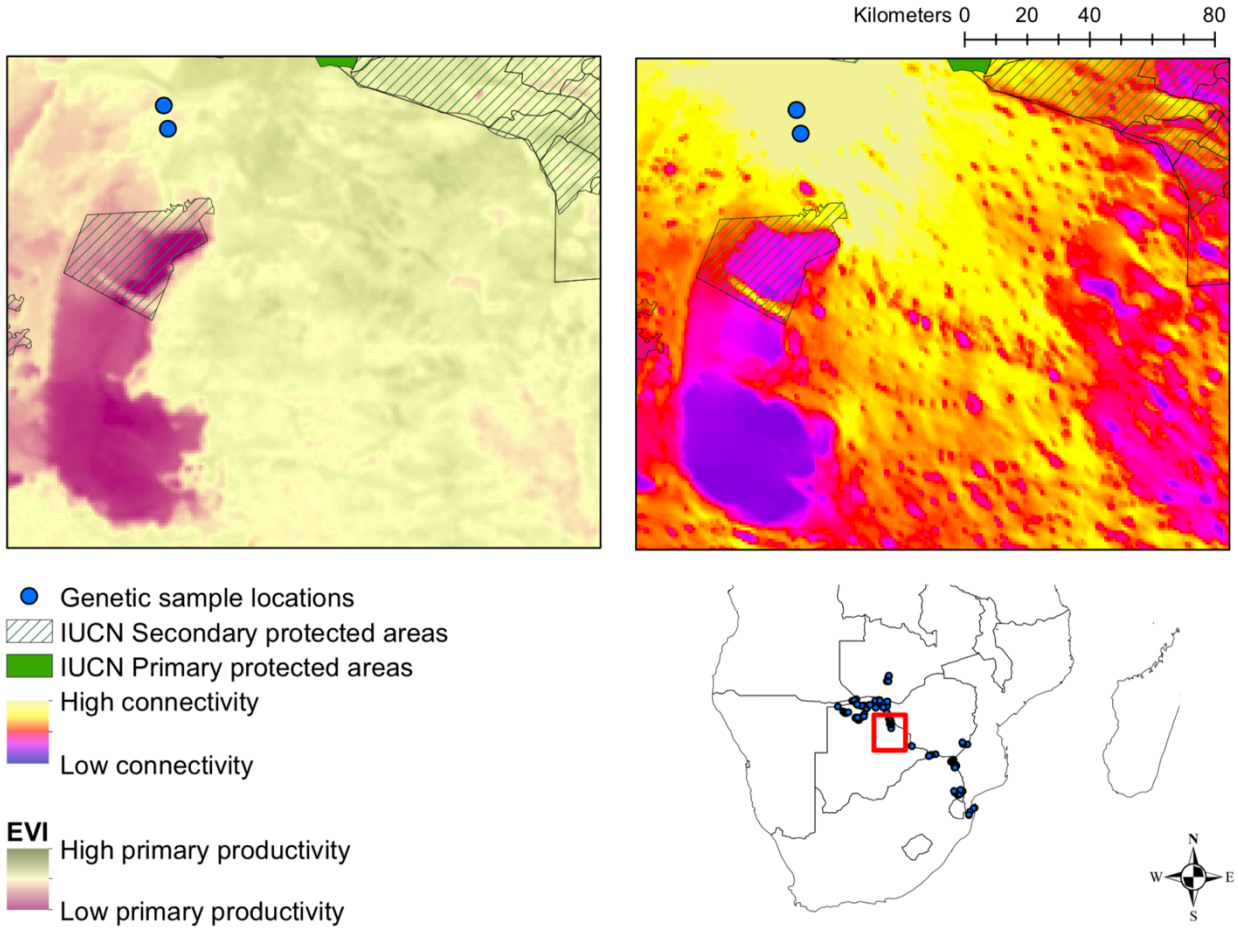

Figure A.9 There are areas in our FLL map where primary productivity as a proxy of food availability seems to be the primary driver of connectivity. For example, areas of high low primary productivity (purple in the left panel) correspond to areas with low landscape connectivity (purple in the right panel).

APPENDIX B: SUPPLEMENTARY TABLES FOR CHAPTER 3

Table B.1 Marker names, profiles and characteristics[4] of nine microsatellite loci used in this study.

Marker name	Locus name	Repeat motif	Size (bp)	Ta-°C	Number of alleles	Ho	He
LaT08[2]	1	(TAGA)16	166–234	56	13	0.774	0.861
Lat13[2]	2	(CATC)21	234–262	56	7	0.650	0.769
Lat17[2]	3	(GGAT)15… (GGAT)	323–355	56	6	0.737	0.779
Lat24[1]	4	(GGAT)22	211–231	56	8	0.694	0.838
FH1[1]	5	(CA)12	81	55	5	0.612	0.661
FH39[1]	6	(CA)18	242	60	11	0.534	0.777
FH102[1]	7	(CT)11(CA)14	179	60	5	0.544	0.557
LA5[3]	8	(CA)13	130–154	52	6	0.576	0.566
Lat25[1]	9	(CCAT)15	298–318	52	7	0.712	0.815
Mean					7.556	0.648	0.736
s.d.					2.744	0.087	0.113

[1] (Comstock, Wasser, and Ostrander 2000)
[2] (Archie, Moss, and Alberts 2003)
[3] (Eggert et al. 2000)
[4] Number of alleles per locus, observed heterozygosity (He) and expected heterozygosity (He)

Table B.2 A list of individuals carrying each of the 28 distinct mitochondrial DNA haplotypes.

Haplotype	N	Genbank Accession Numbers	Individual labels
Hap_1	1	MF062095	1KF05A
Hap_2	5	MF062096	SKNP08A TEP01A SKNP11A SKNP12A TEP02A
Hap_3	4	MF062097	BZ08A BZ16A BZ17A MAP06A
Hap_4	1	MF062098	NKNP19A
Hap_5	2	MF062099	1PT01A 2MR07B
Hap_6	1	MF062100	MAP03A
Hap_7	1	MF062101	1HW01A
Hap_8	4	MF062102	BZ12A BZ13A NKNP02A SKNP07A
Hap_9	1	MF062103	1HW02A
Hap_10	1	MF062104	BZ05A
Hap_11	1	MF062105	BZ19A
Hap_12	3	MF062106	BMAP02A BZ18A SKNP06A
Hap_13	1	MF062107	BZIM02A
Hap_14	18	MF062108	BMAP01A BMAP03A BZ01A BZ04A BZ06A BZ07A BZ10A BZ11A BZ14A BZ15A BZ20A BZ22A NKNP16A NKNP20A NKNP22A NKNP23A SKNP02A SKNP10A
Hap_15	1	MF062109	1KF07A
Hap_16	1	MF062110	1MR13B
Hap_17	3	MF062111	1CP01A 2SS06A GR03A
Hap_18	5	MF062112	1SN13B 2SN05A GR01A GR02A GR05A
Hap_19	4	MF062113	1CH12A 1CP05A 1MR05A 2CH03A
Hap_20	1	MF062114	2SR05A
Hap_21	61	MF062115	1CH02A 1CH07A 1CH13B 1CP04A 1CP06A 1KF01A 1KF08B 1LV02B 1MR01A 1MR02B 1MR03B 1MR04A 1MR07B 1MR08A 1MR10A 1MR11A 1SN03A 1SN11A 1SR01A 1SR03A 1SR10A 2CP02B 2CP04A 2CP05A 2KA01A 2KA03A 2KA04B 2KA05A 2MR01A 2MR04A 2MR06A 2MR08A 2MR10A 2MR12A 2MR13A 2LY01A 2LY03B 2LY05A 2LY08A 2SN01A 2SN03B 2SN07A 2SN11A 2SR01A 2SR02B 2SR07A 2SR08B 2SR09A 2SR10A 2SR11A 2SS03B 2SS04A 2SS05A 2SS07B 2SS09A 2SS10B 2SS11A LMAP01A MAP02A NKNP01A NKNP10A
Hap_22	20	MF062116	1LV01A 1MR06A 1MR14A 1SN01B 1SN05A 1SN09A 1SN10B 1SN12A 2CH01A 2CH02A 2CP03A 2KA02B 2MR11B 2SN02A GR04A MAP01A NKNP05A NKNP06A NKNP09A NKNP11A
Hap_23	1	MF062117	2SR03A
Hap_24	1	AJ428946.1	
Hap_25	2	EF588275.2 DQ316068.1	
Hap_26	1	EU153451.1	
Hap_27	1	DQ188829.2	
Hap_28	1	JF912200	

N = number of individuals carrying that haplotype. Sampling locality abbreviations: KF = collected from Kafue National Park, Zambia; SKNP and NKNP = southern and northern Kruger National Park, South Africa, respectively; TEP and BMAP = Tembe Elephant Reserve and Maputo Elephant Reserve, Mozambique; BZ = the border of Botswana and Zimbabwe; MAP, LMAP and BZIM = Mapungubwe National Park and along the Limpopo river that connects Mapungubwe and Kruger National Park, South Africa; PT = Pandamatenga, Botswana; MR = Moremi Game Reserve, Botswana; HW = Hwange National Park, Zimbabwe; CP = Caprivi region, Namibia; SS and SN = south and north of Savuti Game reserve, Botswana, respectively; GR = Gonarezhou National Park, Zimbabwe; CH = Chobe National Park, Botswana; SR = Seronga, Botswana; LV = Livingstone, Zambia; LY= Linyanti River, Botswana. Labels for Haplotype 24-28 refer to Genbank Accession numbers of outgroup taxa.

Table B.3 Number of elephants and haplotypes of elephants found for each of the mtDNA haplotype clusters identified by Geneland. Number of individuals that carry a haplotype are listed in brackets after the haplotype name.

Cluster Name	Number of individuals	mtDNA Haplotypes
mtDNA Cluster 1	10	Hap_2(5); Hap_3(4); Hap_4(1)
mtDNA Cluster 2	97	Hap_15(1); Hap_16(1); Hap_17(3); Hap_18(5); Hap_19(4); Hap_20(1); Hap_21(61); Hap_22(20); Hap_23(1)
mtDNA Cluster 3	33	Hap_5(2); Hap_6(1); Hap_7(1); Hap_8(4); Hap_9(1); Hap_10(1); Hap_11(1); Hap_12(3);Hap_13(1); Hap_14(18)
mtDNA Cluster4	1	Hap_1(1)

Table B.4 Sample size and the mean (± standard error) number of alleles, number of effective alleles, observed heterozygosity, and expected heterozygosity for the different sampling regions in the study area.

	N	Na	Ne	Ho	He
CNP	6	4.000 (±0.408)	3.124 (±0.374)	0.602 (±0.085)	0.644 (±0.039)
HNP	4	3.778 (±0.278)	3.091 (±0.312)	0.741 (±0.084)	0.651 (±0.033)
LNP	2	2.000 (±0.289)	1.889 (±0.242)	0.389 (±0.139)	0.375 (±0.095)
BZB	13	6.333 (±0.624)	4.243 (±0.446)	0.767 (±0.052)	0.740 (±0.029)
GNP	5	4.444 (±0.294)	3.517 (±0.306)	0.661 (±0.073)	0.694 (±0.033)
MNP	6	3.444 (±0.530)	2.439 (±0.354)	0.504 (±0.101)	0.493 (±0.090)
KNP	33	6.444 (±0.784)	3.956 (±0.453)	0.626 (±0.036)	0.712 (±0.040)

N = sample size, Na = mean number of alleles, Ne number of effective alleles, Ho = observed heterozygosity, He = expected Heterozygosity. Chobe National Park (CNP) = CH, MR, CP, SS, SN, LY; Hwange National Park (HNP) = HW, PT; Livingstone National Park (LNP) = LV; Botswana-Zimbabwe border (BZB) = BZ; Gonarezhou National Park (GNP) = GR, Mapungubwe National Park (MP) = MAP, LMAP, BZIM; Kruger National Park (KR) = NKNP, SKNP

Table B.5 Analysis of molecular variance (AMOVA) of seven nDNA regions showed low differentiation between nDNA regions (Fst), with most of the variation attributed to differences within regions.

Source of variation	**df**	**Sum of squares**	**Estimated variance**	**Percentage of variation**
Among clusters	6	29.457	0.094	3
Within clusters	131	444.456	3.370	97
Total	137	470.913	3.461	100
Fixation index (Fst):	0.027			

Table B.6 There was no evidence for a recent reduction in effective population size assuming a stepwise mutation model (SMM) or a two-phase model (TPM) for Sign and Wilcoxon test. Prob(H>He) is the probability that the heterozygosity (H) is larger than the average (He) under the null hypothesis, if Prob(H>He) is lower than 0.05, the null hypothesis (mutation drift equilibrium) is rejected in favor of the hypothesis of a recent genetic bottleneck (Cornuet and Luikart 1996).

Test	Model	Prob(H>He)
Sign Test	SMM	0.536414
	TPM	0.525722
Wilcoxon Test (1 tailed)	SMM	0.714844
	TPM	0.455078
Wilcoxon Test (2 tailed)	SSM	0.652344
	TPM	0.910156

Table B.7 Genetic diversity indices for all individuals combined (All) and only individuals found in Kruger National Park (Kruger)

Locus	Number of alleles		Expected heterozygosity	
	All	Kruger	All	Kruger
1	13	11	0.86140	0.84997
2	7	6	0.76947	0.76215
3	6	6	0.77907	0.78312
4	9	7	0.83819	0.82073
5	5	5	0.66143	0.70933
6	11	9	0.77676	0.74576
7	5	3	0.55784	0.51329
8	6	5	0.56604	0.52429
9	7	6	0.81535	0.80995
Mean	7.667	6.444	0.73617	0.72429
s.d.	2.784	2.351	0.11362	0.12376

Table B.8 Sample collection and import was sanctioned the Department of Wildlife and National Parks (Botswana), South African National Parks (SANParks) and the Department of Agriculture, Forestry and Fisheries (DAFF), South Africa and by the Zambian Wildlife Authority. In accordance with the Animal Diseases Act, 1984 (Act 35 of 1984) DAFF permits allowed for the import of elephant dung samples from Botswana, Mozambique, Namibia and Zambia, and a SANParks removal permit allowed the removal of samples from Kruger National Park. Permission was granted to collect samples in Kruger National Park in compliance with section 4(1) of the National Environmental Management: Protected Areas Act 57 of 2003. Original permits are available upon request.

Sanctioning Authority	**Permit Description**	**Permit No:**
South African National Parks	Permit to collect natural resources material in the Kruger National Park	SK071
Department of Agriculture, Forestry and Fisheries, South Africa	Veterinary import permit for preserved animal material. Botswana to South Africa. Port of entry Groblersbrug.	13/1/1/30/0-2014/05/001571 and 13/1/1/28/2/10/2-1739
Department of Agriculture, Forestry and Fisheries, South Africa	Veterinary import permit for preserved animal material. Zambia via Zimbabwe to South Africa. Port of entry Beit Bridge.	13/1/1/28/29/8-1742
Department of Agriculture, Forestry and Fisheries, South Africa	Veterinary import permit for preserved animal material. Mozambique to South Africa. Port of entry Lebombo.	13/1/1/28/29/8-1741
Department of Agriculture, Forestry and Fisheries, South Africa	Veterinary import permit for preserved animal material. Namibia to South Africa. Port of entry Nakop/Vioolsdrift.	13/1/1/28/2/10/3-1740
United States Department of Agriculture, United States of America	United States veterinary permit for the importation and transportation of controlled materials and organisms and vectors – Elephant fecal samples	128049 Research

APPENDIX C: SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 4

Table C.1 Identification of the representative sequences from each of the eight mtDNA subclades reported by Ishida et al. (2013) to which the shipwreck ivory mtDNA was compared

GenBank accession number	Sample name	mtDNA subclade identity (Ishida *et al.* 2013)
JQ438737	WA4012	Northern savanna mtDNA subclade
JQ438322	GR0023	Northern savanna mtDNA subclade
JQ438214	BE4059	Northern savanna mtDNA subclade
JQ438674	SW0907	Southeast savanna mtDNA subclade
JQ438459	KR0014	Southeast savanna mtDNA subclade
JQ438139	AM0004	Southeast savanna mtDNA subclade
JQ438649	SE2051	East central mtDNA subclade
JQ438608	NG2182	East central mtDNA subclade
JQ438316	GR0015	East central mtDNA subclade
JQ438637	SA1008	South central mtDNA subclade
JQ438588	NA4704	South central mtDNA subclade
JQ438382	HW0151	South central mtDNA subclade
JQ438503	LO3502	West central mtDNA subclade
JQ438206	BE4035	West central mtDNA subclade
JQ438745	WA4023	West central mtDNA subclade
JQ438501	SL0001	Western mtDNA subclade
AY741327	AY741079	Western mtDNA subclade
AY741079	AY741079	Western mtDNA subclade
JQ438125	AB4527	Savanna wide mtDNA subclade
JQ438407	KE4546	Savanna wide mtDNA subclade
JQ438220	CH0883	Savanna wide mtDNA subclade
JQ438329	GR0038	North central mtDNA subclade
JQ438260	DS1501	North central mtDNA subclade
JQ438208	BE4037	North central mtDNA subclade

Table C.2 Previously reported (Johnson *et al.* 2007) mtDNA sequences corresponding to the Western and West-Central mtDNA subclades, for which the geographic provenance of the elephant had been reported. These were used in a network with sequences generated from shipwreck ivory samples.

GenBank Accession Number	Geographic location	Country	Region/park
AY359266[1]	Central Africa	DRC[2]	
AY359265[3]	Central Africa	Gabon	Aloombe Coast
AY359278[3]	Central Africa	Gabon	Lope Reserve
AF527643[4]	West Africa	Ghana	Kakum National Park
AF527641[4]	West Africa	Ghana	Red Volta Valley
AF527676[4]	West Africa	Ghana	Mole National Park
AF527645[4]	West Africa	Ghana	Kakum National Park
AF527678[4]	West Africa	Ghana	Bia National Park
AF527679[4]	West Africa	Ghana	Bia National Park
AF527642[4]	West Africa	Ghana	Red Volta Valley
AF527675[4]	West Africa	Ghana	Mole National Park
AF527667[4]	West Africa	Mali	Gourma Region
AF527683[4]	West Africa	Ghana	Mole National Park
AF527672[4]	West Africa	Cote d'Ivoire	Tai National Park
AF527673[4]	West Africa	Cote d'Ivoire	Tai National Park
AF527669[4]	West Africa	Mali	Gourma Region
EU096116[5]	Central Africa	CAR[6], Gabon	
EU096117[5]	Central Africa	CAR	
EU096119[5]	Central Africa	Republic of Congo, Gabon	
EU096120[5]	Central Africa	Republic of Congo, Gabon	
EU096121[5]	Central Africa	Gabon	
EU096123[5]	Central Africa	Gabon	
EU096126[5]	Central Africa	Gabon	
EU096128[5]	Central Africa	Gabon	
EU096129[5]	Central Africa	Gabon	
AF106243[7]	West Africa	Ghana	
AF106242[7]	West Africa	Ghana	
AF106245[7]	West Africa	Ghana	
AF106244[7]	West Africa	Ghana	
AY741079[1]	West Africa	Liberia	
AY741327[1]	West Africa	Cote d'Ivoire	
AF527677[3]	West Africa	Ghana	Bia National Park
AF527680[3]	West Africa	Ghana	Bia National Park

Table C.2 (cont.)

AF527674[3]	West Africa	Ghana	Mole National Park
AF527668[3]	West Africa	Mali	Gourma region
AF527670[3]	West Africa	Cote d'Ivoire	Tai National Park
AF527671[3]	West Africa	Cote d'Ivoire	Tai National Park

[1]Debruyne (2005)
[2]Democratic Republic of the Congo
[3]Debruyne et al. (2003)
[4]Eggert et al. (2002)
[5]Johnson et al. (2007)
[6]Central African Republic
[7]Nyakaana et al. (2002)

Table C.3 Newly sequenced twentieth century reference samples from the collection at the University of Cape Town including elephants from Angola, Benin, Burkina Faso, Liberia, Niger and Sierra Leone.

Catalog number	Type of sample	Geographic origin
UCT 16396	Ivory	Angola
UCT 16403	Ivory	Angola
UCT 16402	Ivory	Angola
UCT 16400	Ivory	Angola
UCT 16401	Ivory	Angola
UCT16399	Ivory	Angola
UCT 16422	Ivory/bone	Burkina Faso
UCT 16425	Ivory/bone	Parc W, Niger
UCT 16411	Ivory/bone	Parc W, Niger
UCT16419	Ivory/bone	Sapo National Park, Liberia
UCT 16417	Ivory/bone	Sapo National Park, Liberia
UCT 16428	Ivory/bone	Sapo National Park, Liberia
UCT 16416	Ivory/bone	Sapo National Park, Liberia
UCT 18745	Molar	Pendjari National Park, Benin
UCT 18744	Molar	Pendjari National Park, Benin
UCT 16421	Ivory/bone	Gola Forest, Sierra Leone

Table C.4 Nine shipwreck ivory samples for which mitogenomes were assembled. The complete reference African forest elephant mitogenome consists of 16,109 bp (Brandt *et al.* 2012b).

Sample name	Subclade	Average coverage across mitogenome (X)	Nucleotides (bp) assembly length
B6130	Western	36.54 X	16109 bp
B6074	Western	5.62 X	14970 bp
B6059	Western	8.78 X	15901 bp
B8030	Western	24.68 X	16108 bp
B6079	Western	24.93 X	16097 bp
B6022	Western	7.04 X	14986 bp
B6044	West-Central	4.8 X	15544 bp
B6522	West-Central	3.24 X	14909 bp
B6528	West-Central	3.1 X	14763 bp
Average for all mitogenomes:		13.19 X	15487 bp

Table C.5 Previously published mitogenome sequences used to generate a phylogeny that also included mitogenomes from nine shipwreck ivory samples (SI Table 4).

GenBank accession number	Geographic location	Species	Source
KY616976	Central African Republic: Dzanga Sangha	*Loxodonta cyclotis*	Meyer et al. (2017)
KY616979	Gabon: Lope National Park	*Loxodonta cyclotis*	Meyer et al. (2017)
KY616978	Gabon: Lope National Park	*Loxodonta cyclotis*	Meyer et al. (2017)
KJ557423	Gabon: Lope National Park	*Loxodonta cyclotis*	Finch et al. (2014)
KJ557424	Cote d'Ivoire: Tai NP	*Loxodonta cyclotis*	Finch et al. (2014)
JN673263	Central African Republic: Dzanga Sangha	*Loxodonta cyclotis*	Brandt et al. (2012)
JN673264	Sierra Leone	*Loxodonta cyclotis*	Brandt et al. (2012)
AJ224821	No location information	*Loxodonta africana*	Hauf et al. (2000)
AB443879	No location information	*Loxodonta africana*	Murata et al. (2009)
EU155210	Siberia	*Mammuthus primigenius*	Gilbert et al. (2008)
EU153449	Siberia	*Mammuthus primigenius*	Gilbert et al. (2008)

Table C.6 Species diagnostic single nucleotide polymorphisms (SNPs) in regions of the nuclear genes *BGN*, *PHK*, and *PLP* which were sequenced in shipwreck ivory samples matched SNP character states found in African forest elephants (*Loxodonta cyclotis*). Species diagnostic SNPs are described by Ishida et al. (2011).

	BGN-s2								**PHK A2-s1**		**PHK A2-s2**		**PLP -s1**		
ID[1]	472	485	499	508	513	515	516	570	39	71	871	872	319	345	361
Loxodonta africana	A	T	D[2]/T	G	G	G/D	D/G	C	T	A	T	G	G	T	A
Loxodonta cyclotis	A/G	C	D	T	G	D	D	T	C	A	C	G/A	G/A	T	G
Elephas maximus	A	C	T	G	A	G	D	T	C	G	C	G	G	C	G
Ancient Ivory	A	C	D	T	G	D	D	T	C	A	C	A	G/A	T	G

[1]Reported by Ishida et al. (2011).
[2]D indicates a deletion
[3]When considering unlinked chromosome segments, we were able to amplify a set of eight SNPs for 12 individuals for the BGN-s2 gene region (total: 96 SNPs), a set of two SNPs for PHKA2-s1 for 18 individuals and one set of two SNPs for PHKA2-s2 for two individuals for the PHKA2 gene region, and a set of three SNPs for 8 individuals for the PLP-s1 gene region. A total of 24 SNP loci on 3 genes (BGN, PHKA2, PLP) were examined. For DNA sequences that differentiate between African forest and savanna elephants, DNA from each of the sequenced ancient ivory samples matched the character states found in the African forest elephant, while none of the ancient ivory sequences matched character states unique to the African savanna elephant.

Table C.7 Shipwreck ivory UCT laboratory number, National Museum of Namibia sample number, $\delta^{15}N$ and $\delta^{13}C$ values as well as collagen quality indicators: %N, %C and C:N (atomic) ratios. Collagen extracts in this study had C:N ratios ranging from 3.1 to 3.4, with %C (by weight) from 33.5 to 44.7 and %N from 11.8 to 16.1. C:N ratios between 2.9 and 3.6 indicate well preserved collagen (Ambrose 1990; Van Klinken 1999). In the last column, "x" indicates samples for which mtDNA was analyzed.

UCT No.	Sample No.	$\delta^{15}N$(‰)	$\delta^{13}C$(‰)	%N	%C	C:N	mtDNA
15387	B 6116	7.4	-21.1	14.3	39.1	3.2	
15388	B 6059	7.9	-19.0	15.0	41.1	3.2	x
15389	B 6522	7.1	-21.0	15.1	41.5	3.2	x
15390	B 6127	5.4	-18.7	14.5	39.5	3.2	x
15391	B 6026	7.2	-17.9	15.1	41.7	3.2	
15392	B 6056	7.8	-21.0	14.7	40.2	3.2	x
15393	B 6123	7.9	-17.7	14.1	38.3	3.2	
15394	B 6074	7.2	-17.4	14.1	38.4	3.2	x
15395	B 6019	7.3	-21.1	14.8	40.2	3.2	x
15396	B 6525	7.1	-20.1	14.1	38.1	3.2	x
15583	B 6014	7.5	-18.7	16.0	43.5	3.2	x
15584	B 6055	7.6	-20.1	16.1	44.7	3.2	x
15585	B 6053	7.1	-21.6	16.0	43.5	3.2	x
15586	B 6528	6.8	-20.9	13.5	37.1	3.2	x
15587	B 6022	6.0	-22.0	13.4	36.8	3.2	x
15588	B 6531	7.3	-17.9	15.9	43.2	3.2	x
15589	B 6020	5.6	-22.1	14.8	40.5	3.2	
15590	B 6086	7.4	-21.1	15.9	42.9	3.2	x
15591	B 6028	6.8	-19.9	16.0	43.3	3.2	x
15592	B 8030	6.6	-19.7	15.4	42.0	3.2	x
16088	B 8303	7.7	-20.1	14.1	38.6	3.2	x
16089	B 6085	6.7	-19.6	15.8	43.0	3.2	
16090	B 6129	5.6	-20.9	15.9	43.6	3.2	
16091	B 6073	7.2	-20.2	14.3	39.2	3.2	
16092	B 6044	6.8	-21.2	15.6	43.1	3.2	
16093	B 6078	6.9	-20.2	15.8	43.6	3.2	x
16094	B 8033	6.8	-20.7	15.8	43.4	3.2	
16095	B 6025	7.0	-22.1	15.8	43.6	3.3	x
16096	B 6047	6.0	-18.8	15.7	43.6	3.3	x
16097	B 6082	8.5	-20.5	15.0	41.2	3.2	
16238	B 8043	6.3	-21.3	13.9	38.0	3.2	
16239	B 6031	6.4	-21.9	15.1	40.9	3.2	
16240	B 6012	7.2	-20.8	13.5	36.8	3.2	

Table C.7 (cont.)

16241	B 6032	5.9	-21.6	15.1	40.8	3.1	
16242	B 6065	7.6	-19.7	15.0	42.0	3.3	
16243	B 6057	7.2	-19.9	15.7	42.9	3.2	
16244	B 6033	5.6	-20.5	15.8	43.0	3.2	x
16245	B 6130	6.7	-19.6	15.9	43.5	3.2	x
16246	B 6070	8.6	-21.8	15.5	42.4	3.2	
16247	B 6072	7.2	-21.3	15.0	40.3	3.1	
16248	B 6530	7.4	-21.4	16.0	43.6	3.2	x
16249	B 6122	6.1	-21.6	15.1	41.2	3.2	
16250	B 6136	7.2	-19.3	15.3	41.9	3.2	x
16251	B 6520	6.8	-20.0	15.1	41.0	3.2	x
16252	B 6523	5.1	-21.4	14.9	41.2	3.2	
16253	B 6133	5.0	-18.7	14.8	40.8	3.2	
16254	B 6123	6.8	-20.3	14.0	38.2	3.2	
16255	B 6118	7.6	-21.2	14.8	41.1	3.2	
16256	B 6058	6.7	-21.4	14.7	40.7	3.2	x
16258	B 6062	5.9	-20.6	15.7	42.9	3.2	x
16259	B 6255	6.3	-21.5	13.0	35.5	3.2	
16261	B 6131	8.1	-19.6	14.2	40.4	3.3	
16262	B 6046	6.9	-22.2	15.1	42.9	3.3	
16263	B 6524	6.5	-20.7	15.1	41.7	3.2	
16264	B 6048	7.6	-21.8	15.7	43.1	3.2	
16265	B 6045	6.1	-21.2	15.1	41.0	3.2	
16266	B 6128	6.5	-18.1	15.8	42.9	3.2	x
16267	B 6054	9.0	-19.2	15.8	43.6	3.2	
16268	B 6521	7.0	-20.4	13.2	36.0	3.2	
16271	B 6527	6.9	-20.5	13.6	37.2	3.2	
16272	B 6076	5.8	-19.8	14.8	40.9	3.2	
16274	B 6027	4.8	-21.7	13.1	35.7	3.2	
16275	B 6529	6.3	-19.9	14.2	38.4	3.1	x
16276	B 6526	8.0	-21.6	14.2	38.7	3.2	
16277	B 6087	6.6	-21.9	14.6	39.7	3.2	
16278	B 6023	6.8	-21.0	15.3	41.7	3.2	x
16279	B 6083	7.2	-20.1	15.0	41.2	3.2	x
16280	B 6075	7.4	-19.3	14.1	39.3	3.3	x
16281	B 6077	5.4	-19.9	13.7	37.2	3.2	x
16282	B 6084	7.3	-20.5	14.5	39.4	3.2	
16283	B 6060	5.8	-19.8	15.0	41.4	3.2	x

Table C.7 (cont.)

16284	B 6052	6.4	-20.3	15.0	41.2	3.2	
16286	B 6068	5.2	-18.8	14.4	40.1	3.2	x
16287	B 6067	6.7	-19.7	14.6	39.9	3.2	
16290	B 6527	7.3	-19.4	15.2	41.3	3.2	
16291	B 6064	6.2	-19.5	14.7	40.8	3.2	x
16292	B 6029	6.4	-21.9	11.8	33.5	3.3	
16293	B 6088	6.0	-21.0	14.5	40.1	3.2	x
16294	B 6061	5.8	-20.3	15.2	42.1	3.2	x
16296	B 6066	6.0	-16.9	15.5	42.1	3.2	
16333	B 6079	7.3	-21.3	15.4	42.1	3.2	x
16334	B 6030	6.8	-18.6	15.2	41.3	3.2	x
16335	B 6016	6.3	-20.7	14.6	39.8	3.2	
16336	B 6024	6.7	-20.9	14.5	39.5	3.2	
16337	B 6071	7.2	-17.6	15.9	43.2	3.2	x
16338	B 6081	7.4	-19.6	14.0	38.5	3.2	
16339	B 6069	7.0	-20.0	15.1	40.7	3.1	
16340	B 6021	7.3	-20.1	14.0	38.3	3.2	x
16341	B 6532	7.9	-19.3	14.6	39.3	3.1	
16342	B 6063	8.0	-22.0	14.1	38.4	3.2	
16343	B 6049	7.7	-19.8	13.8	37.6	3.2	
16344	B 6132	5.9	-21.6	15.1	41.2	3.2	
16345	B 6017	7.0	-20.2	14.7	39.5	3.1	
16346	B 6051	7.0	-20.9	14.7	39.8	3.2	x
16347	B 6050	7.4	-20.0	15.0	40.7	3.2	
16348	B 6015	6.6	-20.3	14.2	38.9	3.2	x
16356	B 6013	6.3	-20.5	12.1	34.6	3.4	

Table C.8 Unpublished twentieth century reference samples from the University of Cape Town collections, including $\delta^{15}N$ and $\delta^{13}C$ values as well as collagen quality indicators: %N, %C and C:N (atomic) ratios. In the last column, "x" indicates samples for which mtDNA was analyzed.

UCT No.	Location	Tissue	$\delta^{15}N$(‰)	$\delta^{13}C$(‰)	%N	%C	C:N	mtDNA
16399	Angola	ivory	6.2	-21.3	15.8	43.2	3.2	x
16398	Angola	ivory	7.3	-21.1	15.2	42.1	3.2	
16400	Angola	ivory	8.6	-20.9	14.8	42.0	3.3	x
16404	Angola	ivory	8.9	-20.9	14.6	40.7	3.2	
16403	Angola	ivory	9.4	-20.6	15.4	43.1	3.3	x
16405	Angola	ivory	7.9	-20.2	14.6	40.4	3.2	
16401	Angola	ivory	9.7	-19.5	15.0	41.6	3.2	x
16396	Angola	ivory	8.5	-19.4	15.5	42.4	3.2	x
16397	Angola	ivory	9.2	-18.6	14.2	39.7	3.3	
16402	Angola	ivory	11.3	-17.8	14.9	41.4	3.2	x
18744	Benin	molar	6.8	-22.6	15.5	43.0	3.2	x
18745	Benin	molar	6.8	-22.5	15.4	42.8	3.2	x
18748	Benin	molar	6.0	-22.3	15.5	43.0	3.2	
18746	Benin	molar	8.9	-22.3	15.4	43.0	3.2	
18747	Benin	molar	6.6	-21.8	15.4	43.0	3.2	
18027	Benin	molar	5.4	-21.0	15.4	42.0	3.2	
16410	Burkina Faso	bone	4.8	-21.7	15.2	41.9	3.2	
16422	Burkina Faso	bone	5.2	-21.5	15.4	42.3	3.2	x
18024	Chad	molar	7.3	-21.0	14.7	40.2	3.2	
18023	Chad	molar	9.7	-20.8	15.2	41.6	3.2	
18025	Chad	molar	8.8	-20.7	15.0	41.1	3.2	
18022	Chad	molar	7.8	-20.6	15.4	42.1	3.2	
18021	Chad	molar	8.1	-20.5	15.3	41.5	3.2	
18026	Chad	molar	8.5	-20.4	15.0	41.0	3.2	
16423	Liberia	bone	7.1	-27.2	16.0	43.9	3.2	
16428	Liberia	bone	7.2	-27.1	14.8	44.8	3.5	x
16419	Liberia	bone	7.4	-27.1	16.4	45.1	3.2	x
16416	Liberia	bone	7.4	-27.1	15.3	42.4	3.2	x
16426	Liberia	bone	7.6	-27.1	15.8	43.6	3.2	
16417	Liberia	bone	7.1	-26.5	15.7	43.2	3.2	x
16412	Liberia	bone	7.5	-26.4	15.9	43.2	3.2	
16414	Liberia	bone	8.3	-26.2	14.7	40.1	3.2	
16413	Liberia	bone	8.0	-26.2	15.9	43.2	3.2	

Table C.8 (cont.)

16424	Malawi	bone	6.3	-19.9	15.4	42.8	3.2	
16431	Malawi	bone	6.2	-19.6	16.0	44.0	3.2	
16411	Niger	bone	5.2	-21.8	16.0	43.7	3.2	x
16425	Niger	bone	5.3	-21.8	15.1	42.5	3.3	x
16430	Niger	bone	5.3	-21.4	15.5	42.7	3.2	
16418	Niger	bone	5.3	-21.3	15.9	43.1	3.2	
16421	Sierra Leone	bone	8.4	-24.4	15.5	42.4	3.2	x
16432	South Africa	bone	12.7	-17.2	12.5	34.4	3.2	
16406	South Africa	ivory	8.2	-22.0	15.3	42.0	3.2	
16407	South Africa	ivory	8.4	-19.7	14.8	40.7	3.2	
16409	South Africa	ivory	8.3	-19.1	14.8	40.4	3.2	
16408	South Africa	ivory	7.4	-20.2	16.0	44.0	3.2	
16429	South Africa	bone	13.1	-16.9	15.5	42.8	3.2	
16433	South Africa	bone	8.5	-19.2	15.7	43.7	3.3	
16434	South Africa	bone	9.5	-18.6	15.4	42.1	3.2	
16435	South Africa	bone	8.2	-19.5	15.5	41.7	3.2	
16436	South Africa	bone	7.8	-20.4	15.8	43.3	3.2	
16437	South Africa	bone	9.6	-18.3	15.7	42.5	3.2	

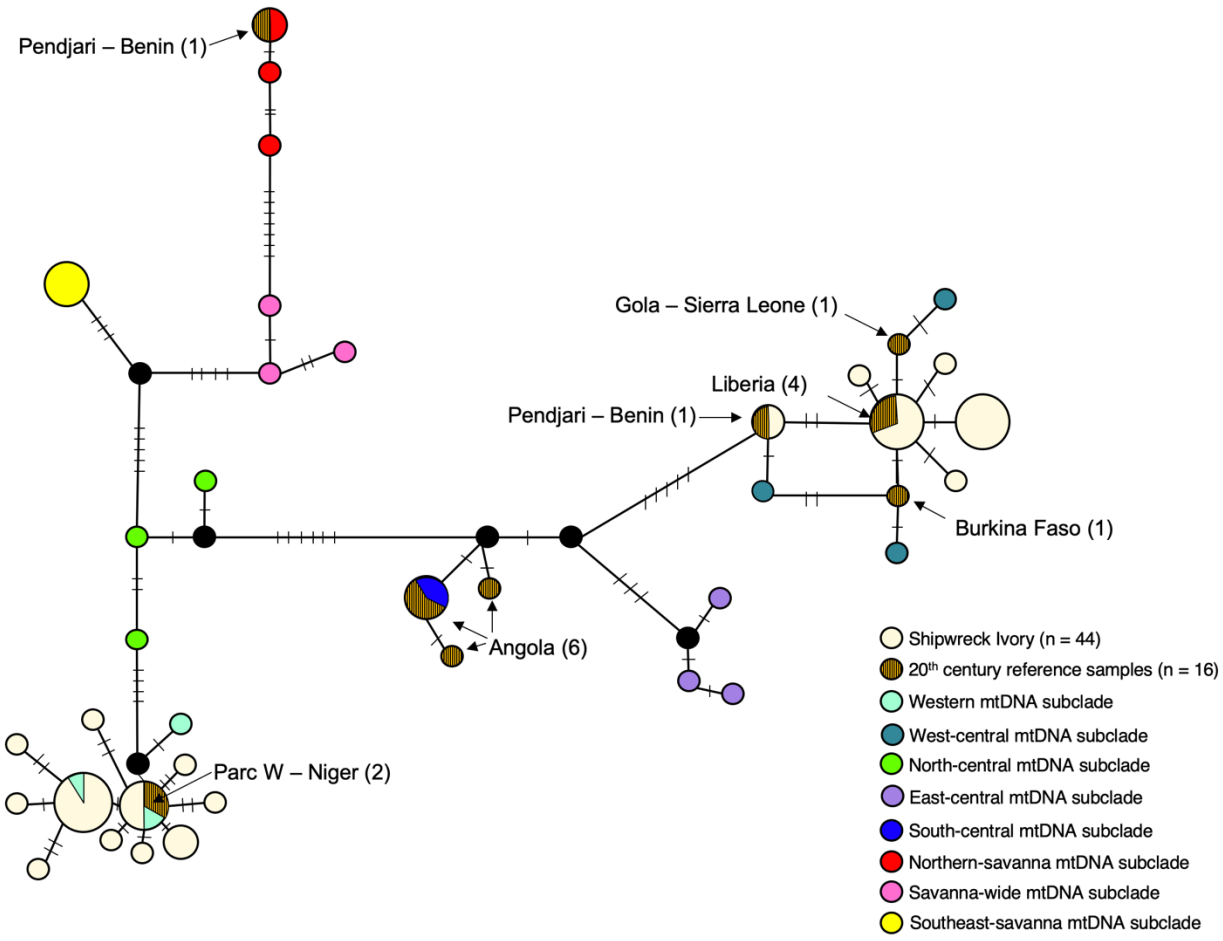

Figure C.1 A median-joining haplotype network (based on pairwise distances between haplotypes) compares shipwreck ivory mtDNA reference sequences (ivory colored) to newly sequenced 20th century reference samples (of known geographic origin) from the University of Cape Town collection (brown with black vertical stripes), and the 8 mtDNA subclades reported by Ishida et al. (2013). The shipwreck ivory haplotypes grouped with the haplotypes reported for 20th century elephants from West Africa, and not within mtDNA subclades common among savanna or central African forest elephants. Branch length is proportional to the number of mutational differences (indicated as cross-hatches) between haplotypes, circle size is related to the number of individuals carrying a haplotype.

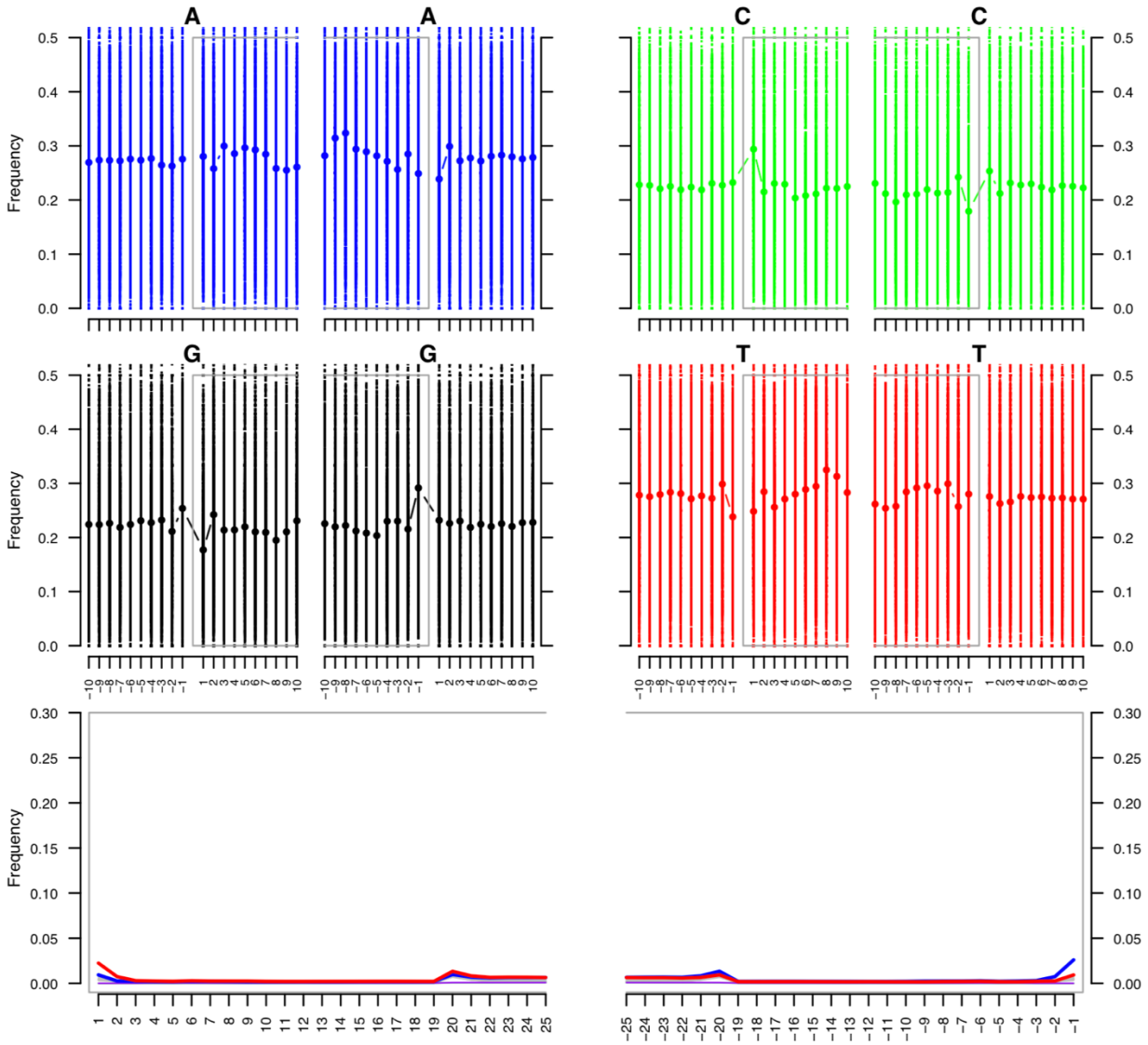

Figure C.2 The shipwreck ivory DNA shows damage patterns characteristic of ancient DNA. The fragment misincorporation plots illustrated above are for a single individual (B6059 belonging to the Western mtDNA subclade – see Figure 1D) and are typical of the damage patterns observed in all of the ancient ivory samples. The smaller top four panels show the base frequency outside and inside the read sequence (the open grey box indicates the read span), and the bottom two plots are the base positions of substitutions from the 5' (left) and the 3' end (right). The bottom plot shows C to T substitutions in red, G to A substitutions in blue, and all other substitutions in grey. This figure was produced using the program mapDamage2 (Jónsson *et al.* 2013).

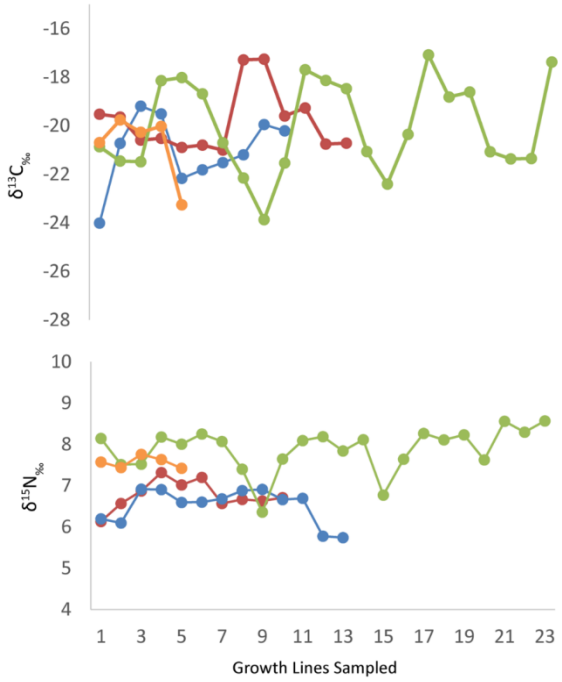

Figure C.3 $\delta^{15}N$ and $\delta^{13}C$ values for four tusks in which successive growth increments were analysed to explore variation in isotope values during the lifetimes of the elephants. Wet/dry season variation in $\delta^{15}N$ is less pronounced than in $\delta^{13}C$, but overall the four tusks do not show significant variation in diet over time (Cerling *et al.* 2009; Codron *et al.* 2012).

APPENDIX D: A STEPWISE PROTOCOL OF HOW TO MODIFY THE R_X SEX IDENTIFICATION SCRIPT

Steps for adapting the Rx_identifier.r script from Mittnik *et al.* (2016) to any target species that has a chromosome level reference genome and XY sex determination.

1. Align the target species sequence reads to the chromosome level reference genome
    a. Convert resulting SAM alignment files to BAM files and filter out unmapped and low-quality reads
    b. Remove PCR duplicates, sort and index SAM files using SAMtools (or similar software)
    c. Generate index statistic files for each BAM file using the "idxstats" function in SAMtools (or similar software). Add the suffix ".idxstats" to each resulting file.
    d. the "idxstats" files are used as input for the Rx_identifier R script. Each row in an "idxstats" file contains the reference sequence name, sequence length, # mapped reads and # unmapped reads.
    e. Verify that idxstats files contain the number of reads mapped to each chromosome and determine which row number represents the X chromosome.
2. Adapt Rx_identifier.r script for the target species of interest
    a. Download the Rx_identifier.r script from Mittnik *et al.* (2016).
    b. Edit line 15 in the script so that nrows reflect the number of chromosomes (autosomal + sex chromosomes) of your target species reference genome (e.g. nrows = 24 for humans, nrows = 28 for African savanna elephants – see Appendix 2, Rx_identifier.r script that has been adapted for elephants).
    c. Rows 24 to 47 calculates for each chromosome the ratio of the alignments to each chromosome to the total number of alignments to autosomes and sex chromosomes
        i. Adjust the script to reflect the number of chromosomes in the target species (i.e. add or remove lines per chromosome so that the final number of lines of script are equal to the number of chromosomes on the target species reference genome).

ii. In the MMittnik *et al.* (2016) script line 23 in the idxstats files represent the X chromosome, therefore Rt23 in the Rx_identifier.r script (line 46 in the script) represents the ratio of alignments to the X chromosome to the total number of alignments to all chromosomes

iii. For the script adapted for elephants (Supplementary Appendix 2), the "idxstats" files contain X chromosome information on the 28th line of each file, therefore Rt28 represents the X chromosome.

d. Edit line 49 in the script so that the numerator represents the ratio associated with the X chromosome. Add or remove fractions so that all chromosomes in the reference sequence are represented in the equation. However, since one is comparing the X chromosome to autosomal chromosomes, do not include the Y chromosome ratio (e.g. Rt24 from line 47 in the Mittnik *et al.* (2016) script) in the calculation.

e. Edit line 52 so that the SE measures the amount of variability in the Rx mean compared with the number of autosomes in the target species (22 in humans, 27 in elephants).

f. Save and add the modified script to the same folder as the "idxstat" files and run the script in R (R version 3.6.1. R Core Team, 2019).

g. For each "idxstats" file the script produces a result file with the suffix ".Rx" which contains information on the sex determination statistics and whether sex identification is possible for the individual.

APPENDIX E: SUPPLEMENTARY TABLES AND FIGURES FOR CHAPTER 5

Table E.1 Examples of organisms for which the reference genome was generated from a female animal or where only the *X* and not the *Y* chromosome is reported in the GenBank reference genome assembly.

Common name	Scientific name	GenBank assembly number
Domestic cat	*Felis catus*	GCA_000181335.4
Domestic sheep	*Ovis aries*	GCA_002742125.1
Domestic horse	*Equus caballus*	GCA_002863925.1
Dromedary camel	*Camelus dromedarius*	GCA_000803125.2
European rabbit	*Oryctolagus cuniculus*	GCA_000003625.1
Domestic goat	*Capra hircus*	GCA_000317765.2
Sumatran orangutan	*Pongo abelii*	GCA_002880775.3
Western lowland gorilla	*Gorilla gorilla gorilla*	GCA_000151905.3
Gelada	*Theropithecus gelada*	GCA_003255815.1
Meerkat	*Suricata suricatta*	GCA_006229205.1

Table E.2 Genome alignment statistics for modern and ancient elephant genomes. Breadth of coverage was calculated as the proportion of the genome covered by at least 1 read. Average coverage was calculated as the depth of coverage across the genome for all positions in the genome, including those with zero coverage.

Sample ID	Sample type	Breadth of coverage (%)	Average coverage (X-fold)
DS1531	Modern	45.4242	0.721645
DS1548	Modern	56.2786	0.982199
DS1514	Modern	39.1834	0.564377
DS1543	Modern	50.7346	0.826475
DS1506	Modern	55.9282	0.969139
LO3503	Modern	52.5948	0.864645
LO3509	Modern	53.8217	0.897441
LO3511	Modern	49.8096	0.791459
LO3521	Modern	47.3061	0.737514
LO3514	Modern	49.4394	0.793958
GR0041	Modern	51.025	0.821681
GR0037	Modern	49.8363	0.806842
WA4013	Modern	12.3095	0.141394
WA4020	Modern	61.1214	1.09633
BE4059	Modern	63.3544	1.16019
B6079	Ancient	7.23345	0.153929
B6025	Ancient	1.4834	0.0300563
B8030	Ancient	7.96246	0.169617
B6059	Ancient	5.23486	0.109229
B6074	Ancient	1.27702	0.0258847
B6022	Ancient	2.69254	0.0550849
B6062	Ancient	1.35766	0.0275851
B6051	Ancient	0.637552	0.0128607
B6130	Ancient	20.6747	0.496995
B6520	Ancient	0.193728	0.00388982

Table E.3 Rx ratio values for the ancient ivory samples calculated using all sequence reads (Rx all), approximately 10 000 sequence reads (Rx ~ 10 000), and approximately 1000 sequence reads (Rx ~ 1000).

Ancient sample ID	Rx all	Rx ~ 10 000	Rx ~ 1000
B6022	0.495573	0.4483153	0.3957278
B6025	0.491787	0.5043557	0.1873106
B6051	0.490834	0.4407088	0.4415854
B6059	0.482732	0.493083	0.3499586
B6062	0.968406	0.9666384	1.029646
B6074	0.491222	0.4447996	0.2836775
B6079	0.477581	0.5970461	0.6939766
B6130	0.474589	0.5681135	0.8121833
B6520	0.491588	0.4147076	0.5607132
B8030	1.00599	0.9682274	0.8207848

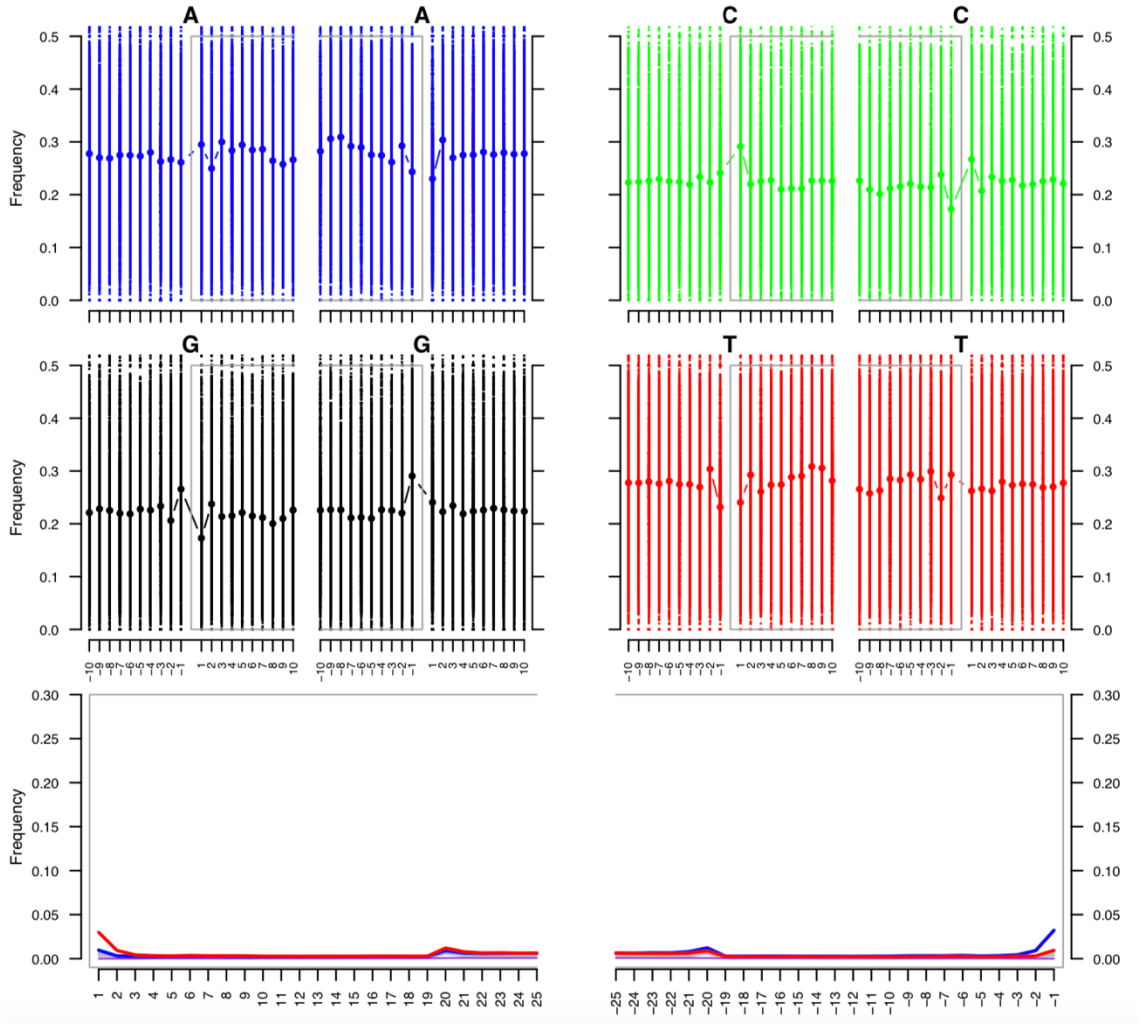

Figure E.1 DNA damage patterns showing fragmentation and misincorporation patterns. The top four plots show base frequency outside and in the read (grey box corresponds to the read), with base positions from the terminal position indicated below the graphs. The bottom plot shows base pair position-specific misincorporations from the 5' (left) and the 3' end (right), red shows C to T substitutions, and blue shows G to A substitutions.

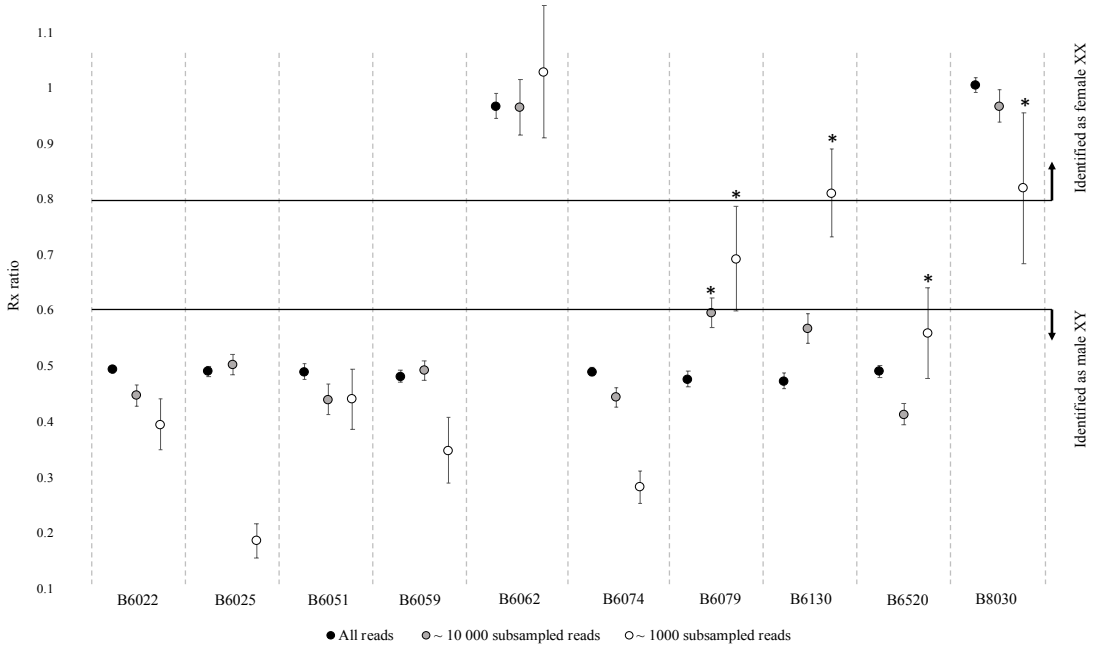

Figure E.2 Rx is the ratio of sequence read alignments to the *X* chromosome compared to sequence read alignments to all autosomes. Rx values are shown for low coverage ancient elephant genomes (black), and for datafiles that were subsampled to include ~10 000 (gray) and ~1000 reads (white). An Rx ratio with an upper 95% CI of less than 0.6 indicates male sex, and an Rx ratio with a lower 95% CI that is greater than 0.8 indicates female sex. For samples that have 95% CI outside of the Rx cut-off values, the Rx method provides the sex identification output as "sample is consistent with XX" or "sample is consistent with XY" rather than definitively scoring the sample as either "male" or "female". This occurred in 5 cases in the subsampled datafiles (indicated with asterisks).